INDIRECT EFFECTS OF FINANCIAL INCENTIVES ON PHYSICIAN BEHAVIOR IN PERINATAL CARE

by

Debra G. Bozzi, MSPH

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

May 2021

Abstract

This dissertation explores how financial incentives affect physician behavior on dimensions of healthcare quality and spending in privately insured markets, by assessing indirect impacts (e.g. effects not directly targeted) of payment reforms for perinatal care. Studies focus on two payment changes to assess effects on: payer-related spillovers, racial disparities, and variation by providers' baseline quality.

Chapter 1 evaluates spillover effects of a 2014 and 2015 Medicaid payment reform in four states that discontinued payment for Early Elective Deliveries (EEDs), a low-value service, on low-value quality measures among commercial enrollees. Results show a small, but significant decrease in EEDs in the commercial sector. There is no evidence of physician-induced demand, under which the rate of more profitable services would increase in response to negative income shocks. Spillovers were concentrated in for-profit hospitals, where financial objectives were aligned between physicians and hospitals. Findings indicate that reducing fees for low-value services can be welfare-improving, through positive impacts on healthcare quality across payers.

Chapter 2 examines the effect of a 2013 multi-payer bundled payment program for perinatal episodes in Arkansas on place-based racial disparities in private insurance markets. This incentive shifted financial risk to physicians by offering a risk-adjusted case rate for the entire episode, rather than reimbursing separately for each service. Results show increased quality improvement in areas with a high proportion of White patients in the short-term; however, areas with a high proportion of Black patients equalized gains in the long-term. This study highlights the role of physician payment on racial disparities, and suggests a need for financial incentives that are directly tied to racial equity to achieve equitable short-term gains.

Chapter 3 leverages an episode-based payment program with compulsory provider participation to test whether baseline quality is associated with divergent impacts on quality and spending. Although areas with low performing providers at baseline achieved larger improvements in quality measures directly linked to payment, areas with high performing providers at baseline saved 16.2% more across episodes, through reduced hospital prices and volume of services. Results suggest that under mandatory reforms, reducing costs in low performing areas may be difficult without compromising quality.

Primary Readers:

Aditi P. Sen, M.A., Ph.D. (Advisor)

Cynthia Minkovitz, M.D., M.P.P. (Chair)

Darrell Gaskin, M.S., Ph.D.

Craig E. Pollack, M.D., M.H.S., M.Sc.

Elizabeth Stuart, Ph.D.

For Buster.

Acknowledgments

This dissertation reflects the support and encouragement of many colleagues, family members, and friends.

A common thread throughout my career has been incredible mentorship. First and foremost, I would like to thank my advisor, Dr. Aditi Sen, for her invaluable guidance. Aditi has helped me become a better researcher and health economist, by teaching me to always think critically about the policy context. This dissertation is defined by her willingness to work with me; knowing she was on my team encouraged me to take risks and pursue opportunities that I may have otherwise passed up. Her mentorship, which has been filled with practical advice, enthusiasm for my work, and a shared love of the city of Philadelphia, has left a lasting impression. I would also like to thank Dr. Craig Pollack, who has given me amazing opportunities to tackle new research questions, and has pushed me to gain confidence as an independent researcher. I am grateful to the remaining members of my thesis advisory committee, Dr. Darrell Gaskin, Dr. Cynthia Minkovitz, and Dr. Elizabeth Stuart. Each member brought a unique and meaningful perspective, offering guidance, suggestions, and contributions to the formulation of this research. There are many other faculty members that I am fortunate to have collaborated with and learned from. Thank you to Dr. Lauren Nicholas for supporting my ability to engage in the methodological, theoretical, and policy-oriented foundations of health economics research, and to Dr. Brad Herring, for teaching me invaluable lessons in developing meaningful research questions and understanding data. Beyond the research, I am grateful for Dr. Matt Eisenberg, Dr. Beth Resnick, and Dr. Jon Vernick, for always advocating for me. I would also like to acknowledge a number of individuals who helped me find the world of health economics. Thank you to my former colleagues at the Pennsylvania Healthcare Quality Alliance and Independence Blue Cross, Erik Muther and Josh Pechar, for helping to spark my interest in health policy and supporting me in the early stages of my career.

I owe special thanks to an important group of friends. I am grateful to my PhD cohort, who made the transition to Baltimore much easier and a lot more fun. I am particularly thankful for Julia Burgdorf, Jon Levin, Molly Merrill-Francis, and Zach Predmore, for their friendship and advice during the dissertation. Thank you to the students who came before me and provided a roadmap for success – especially Alli Oakes, Julia Pringle, and Sonal Parasrampuria – and to my health economics "journal club" pals – Kelly Anderson, Mark Meiselbach, Brendan Rabideau, and Yang Wang. Beyond Hopkins, I am incredibly fortunate to have supportive friends in all walks of life. Specifically, thank you to my friends from North Shore, Haverford, and to those who make my life better in Philly every day.

I am incredibly grateful for my family. Thank you to my in-laws, Catharine and John, for always showing an interest in my work and willingly reading my publications. John even built my desk, an asset that I quite literally could not have finished my dissertation without. Thank you to my brother and sister, Michele and Adam. I am lucky to have a sister in health policy to seek advice from, and a brother who will always drop everything to have fun with me. Thank you to my mom, Lynn, for giving me the safety, security, and confidence to explore my passions from an early age, and to Marty, for being a constant source of reason.

I am indebted to my "coworkers," Frankie and Maya, who always provided the right mix of shenanigans and snuggles. I am especially thankful for Buster, who was my best little buddy for the first four years of the program. Finally, I would not be here without the best husband, Mike. Thank you for encouraging me to pursue the doctoral program, even though it meant living in different cities. You helped make the most of our time off together, planning hikes, baking cookies, and most importantly, making me laugh. Thank you for listening to my presentations more times than you would like, and for being my biggest cheerleader.

It has been an especially unique experience writing my dissertation in the clinical setting of perinatal care while being pregnant, and I am incredibly excited for what the future holds.

Table of Contents

List of Tables

List of Figures

1. Introduction

Physician reimbursement is the financing mechanism under which insurers pay doctors for care. As such, it plays a significant role in the healthcare system, shaping the amount and type of medical care that a physician provides. The traditional payment system in both public and private insurance markets in the U.S. is fee-for-service (FFS), where doctors are paid for each service performed, regardless of the quality or outcome of care. As recently as 2016, 71% of overall physician revenue was derived from FFS payments, a number that has remained fairly constant in the last decade, despite a decrease in the number of practices receiving some FFS income (Rama, 2017). Since physicians are not held financially accountable for their patients' outcomes under this payment system, FFS has incentivized the overprovision of services, leading to an increase in overall healthcare costs over time (Ikegami, 2015). As a share of gross domestic product (GDP), total healthcare spending has more than doubled since 1975, increasing from 7.9 percent to 17.6 percent (MedPAC, 2019). Nonetheless, quality of care is incommensurate with the level of spending, as the U.S. ranks the lowest on health performance indicators among comparable nations (Chee, Ryan, Wasfy, & Borden, 2016).

Government policies have, for decades, introduced new approaches to physician payment, experimenting with public and private sector reforms to balance the dual considerations of beneficiaries' access and quality of care against adequacy of payment rates to support provider costs and access to capital, while also ensuring an efficient use of healthcare services (MedPAC, 2020). This has generally involved moving payment systems away from FFS towards value-based payments that impose financial risk on physicians to generate cost savings and increase quality of care. Still, a longstanding issue of health economics is the optimal design of these financial incentives – how can government policies design physician payment as a means of improving quality of care and the resultant health outcomes, while minimizing healthcare expenditures?

The Affordable Care Act (ACA) accelerated the movement towards value-based care delivery and payment models, while simultaneously reducing the annual increase in payment rates in public insurance markets to generate savings (Blumenthal & Abrams, 2020). Value-based programs are budget-neutral, meaning that providers receiving higher payments do so at the expense of the other participants, introducing concerns that penalties may disproportionately impact physicians serving disadvantaged populations (Gilman, Adams, et al., 2015). This is problematic because vulnerable populations, such as those with higher-than-average medical and social needs, incur high healthcare costs and tend to have poor outcomes (Joynt Maddox, 2018). In 2016, the costliest 5 percent of beneficiaries accounted for 40 percent of annual Medicare spending; in contrast, the least costly 50 percent accounted for only 5 percent of spending (MedPAC, 2019). In five federal value-based programs, providers predominantly serving medically and socially vulnerable patients had an 18% greater probability of being penalized under value-based payment (Joynt Maddox, 2018). In 2016, these physicians and hospitals operated at an average profit margin of 1.6 percent, relative to the national average of 7.8 percent (Greenberg, 2020). Redistributing payment away from providers with the highest resource needs can exacerbate disparities over time and stimulate other downstream effects (Damberg, Elliott, & Ewing, 2015).

In addition to the financial ramifications on physicians serving high-risk patients, financial incentives are also associated with clinical implications. Since the implementation of the ACA, growing financial challenges have led to increased rural and safety net hospital closures, exacerbating overcrowding at neighboring facilities and reducing access to care (Khullar, Song, & Chokshi, 2018; Sachs et al., 2019). In both private and public insurance markets, the financial implications of healthcare payment reform for high-risk patient populations has also led to increased physician gaming. Prior research has found this to include the avoidance of high-risk patients, overprovision of care among beneficiaries in health plans outside of the incentive program, and hesitancy to participate in value-based reimbursement if the financial outlook is uncertain (Mendelson et al., 2017).

The presence of these unintended consequences stems from the notion that efforts to design financial incentives predominantly focus on improving quality for the "average" patient, rather than specific populations. Further, payment reforms lack uniformity across markets and clinical settings, leading to wide variation in incentive magnitude, payment methodology, and cost and quality metrics. This makes financial incentives difficult to study, and also imposes an undue administrative burden on physicians and hospitals (Mendelson et al., 2017). Inattention to indirect impacts – effects not directly targeted by the incentive – across populations and payers can lead to an inadequate understanding of the effects of payment reform, and may hamper the potential to leverage financial incentives as a tool for improving care across populations and settings.

1.1. Economic Framework for Understanding the Role of Financial Incentives on Physician Behavior

The policy debate on how to pay physicians is illustrative of a broader body of economic research on physician agency. Physician agency is the idea that doctors make treatment decisions on behalf of their patients due to asymmetric information; because physicians train extensively to learn appropriate diagnoses and treatments, they serve as medical "experts" to their patients, who lack the knowledge to make autonomous treatment choices (Mcguire, 2000). However, a complicating factor in this relationship is a physician's desire to maximize profits, which can lead to a violation of agency if the patient would not make the same choice if they had all of the same information. This means that the financial incentive a physician receives has potential to lead to an overabundance or shortage of services for the patient, if this behavior raises physician income. To achieve the optimal level of healthcare consumption, insurers have increasingly incorporated financial incentives into reimbursement design that impose financial risk on physicians for healthcare spending and quality of care, to encourage perfect physician agency.

This theory is supported by empirical studies that highlight violations of physician agency, under which physician behavior deviates from traditional economic principles in response to financial incentives. In 1958, Reuben Kessel found evidence of price discrimination, in which physicians scaled the price of care by patient income; wealthier patients faced higher prices to finance charity care for sicker patient who could not afford to pay the average price. This behavior contradicted standard economic principles; typically, profit-maximizing firms eliminate this practice in competitive markets due to inefficiency. Instead, doctors would be expected to establish uniform prices for identical services to stimulate demand among affluent patients. This paper introduced the idea that physicians did not strictly prioritize the maximization of profit; they also sought to improve patient well-being (Kessel, 1958). In 1978, Victor Fuchs observed a direct relationship between an increased supply of surgeons and market price, despite steady patient demand. This result was unexpected because in traditional economics, increased competition is correlated with a reduction in price. This study introduced the notion that physicians have the ability to influence patient demand, independent of the payment rate (Fuchs, 1978). Finally, Thomas Rice's 1983 study observed an increase in the volume of services supplied in the presence of negative income shocks. This challenged a very basic norm in economics, in which supply falls with the price. It gave rise to the existence of physician-induced demand, a type of imperfect physician agency where practitioners prompt patients to consume more than the optimal amount of care to increase profits (Rice, 1983). These studies highlight the unique nature of physician behavior, and the need to better understand how financial incentives put forth by variation in prices impacts the provision of healthcare. Other research has found that cost-increasing losses that arise under FFS can be mitigated by incentives that introduce financial risk bearing to physicians (Mendelson et al., 2017; Town, Wholey, Kralewski, & Dowd, 2004). However, the overall welfare effects depend on the magnitude of the physician response, and variation in effects across payers and patient populations (Baicker, Chernew, & Robbins, 2013).

An ongoing public policy debate is the translation of these ideas into the design of physician reimbursement. In the 1980s, economists Thomas McGuire and Randall Ellis postulated that if physicians undervalue benefits to patients relative to profits, then cost-based reimbursement leads to the overprovision of care, while capitated payment results in a shortage of services for patients (Ellis & McGuire, 1986). This idea, coupled with empirical findings, led to the proliferation of mixed reimbursement policies aimed at concurrently containing costs and improving quality. Examples include the prevalence of pay-for-performance policies in public and private insurance and the design of bundled payments, which shift financial risk to physicians through shared savings and penalties for specific episodes of care. However, more recent evidence has begun to highlight unintended consequences of these approaches, including avoidance of high-risk patients and widening disparities for marginalized patient and provider populations (Joynt Maddox, 2018; Mendelson et al., 2017). These findings underscored a need to shift the design of financial incentives towards holding physicians accountable for specific populations. The most prominent example is the Center for Medicare and Medicaid Services' Direct Contracting Model launched in April 2021, which introduced downside risk to physicians serving patients dually enrolled in Medicare and Medicaid, a population with high medical needs.

While the majority of markets have introduced value-based payment reforms, how these policies affect patients beyond the overall population is poorly understood. This dissertation explores specific types of financial incentive policies to better understand the downstream effects on beneficiaries insured by other payers, racial minorities, and populations with low quality of care.

1.2. Importance of Perinatal Care

Financial incentives have been implemented across many clinical services, but perinatal care is a unique setting to explore their indirect effects. From a medical perspective, perinatal care is clinically salient. In the U.S., childbirth is the top expenditure category for hospital payments made by Medicaid and private insurers, yet maternal mortality is higher than all other developed countries (Carroll, 2017; Truven Health Analytics, 2013). Since 2013, average inpatient costs for maternal and child health have increased by 32 percent, compared to overall hospital spending, which grew by 4.8 percent in the same time frame (Kamal & Cox, 2018; Truven Health Analytics, 2013). Between 2000 and 2014, the rate of deaths due to complications from pregnancy and childbirth worsened by 26.6 percent in the U.S., while the global rate improved by 38 percent (MacDorman, Declercq, Cabral, & Morton, 2016). Over half of U.S. hospitals lag below the national target in quality scores, suggesting significant room for improvement (Consumer Reports, 2017). Vulnerable populations, such as low-income mothers and racial minorities, are disproportionately affected, making it especially urgent to address this issue among these patients (Kozhimannil, Hardeman, & Henning-Smith, 2017).

From an economics perspective, perinatal care also offers an ideal setting for studying financial incentives, since it has features expected to enhance generalizability. First, volume is relatively independent from physician decision-making (e.g. compared to elective arthroplasty, which requires a pre-surgical appointment), increasing the likelihood that physician behavior stems directly from payment changes, rather than confounders that would influence treatment decisions. In other words, the decision to have a baby is usually a family one, and is not typically influenced by the physician, whereas the volume of other services common to studying financial incentives is more dependent on physician discretion. Second, perinatal care has features that have been linked to stronger provider responses, including high variation in quality and costs, and the presence of elective services with practical treatment substitutes (Chou et al., 2006). These unique characteristics make perinatal care a conducive setting to enact financial incentives, in that changes in treatment patterns are anticipated to be driven by supply-side, rather than demand-side, factors (Town et al., 2004). Finally, intrapartum delivery is emergent, and under the federal Emergency Medical Treatment and Labor Act, every U.S. hospital with an emergency room is compelled to treat patients who arrive in labor. This presents a distinct opportunity to explore the quality of care across a range of substitutes. The role of financial incentives is typically assessed in clinical settings that rely on a binary indicator for whether a patient received treatment, so it is difficult to discern whether a lack of treatment is equivalent to subpar treatment.

1.3. Summary of Chapters

The goal of this dissertation is to examine the impact of financial incentives on physician behavior on dimensions of healthcare quality and spending for beneficiaries in private insurance markets. This is done by assessing three indirect impacts (e.g. effects not directly targeted) of payment reforms for perinatal care: spillover effects, racial disparities, and variation by providers' baseline quality.

Payment reform policies have traditionally started at state and local levels. In 2014 and 2015, four states passed a Medicaid policy that discontinued payment for Early Elective Deliveries (EEDs), a low-value mode of childbirth defined as a scheduled, non-medically necessary induction of cesarean section before 39 weeks' gestation. The aim of the policy was to leverage a reduction in the price of low-value care to incentivize physicians to reduce EEDs, which are convenient for the physician and patient to schedule, but are not clinically beneficial. While previous evidence has shown significant reductions in EEDs in the Medicaid population, no studies have examined indirect impact of the policy in the non-Medicaid population (Allen & Grossman, 2019; Dahlen, Mccullough, Fertig, Dowd, & Riley, 2017).

In Chapter 1, this policy is used as a case-study to understand whether effects of lowering the price of a low-value service in Medicaid spillover to privately insured patients. This chapter assesses whether the policy leads to positive spillovers, under which desired physician behavior extends beyond the target population, as well as unintended consequences. Using data from Medicare Hospital Compare and the Truven MarketScan Claims and Encounters Data from 2013 to 2017, difference-in-difference analyses were conducted on measures of low-value care across all payers. Analyses explore whether the Medicaid payment policy prompted physicians to reduce the rate of EEDs in the all-payer population (e.g. a positive spillover), and if this effect was accompanied by an increased provision of a more profitable service to compensate for lost income (e.g. physician-induced demand). The hypothesis was that the Medicaid payment reform would be associated with a payer-related spillover through a reduction in EEDs in the commercial sector, and a welfare-reducing increase in physician-induced demand through low-risk c-sections, which are less time-intensive and more profitable, but riskier than vaginal deliveries.

The results show a small, but significant 3.35% decline in all-payer EEDs in states with the Medicaid payment policy compared to control states, with no evidence of physician-induced demand. One major question is whether observed spillovers are prompted by financial or reputational drivers of physician behavior, such as the type of hospital in which a physician is employed. In general, non-profit and for-profit hospitals have different objective functions, where non-profits aim to maximize profits and quality, whereas for-profits focus predominantly on income. As a result, physicians in for-profit hospitals, driven by financial objectives, may respond more strongly to changes in service profitability. Findings are consistent with this expectation, as analyses show larger spillovers in areas with a higher share of for-profit hospitals. This study finds that the Medicaid payment policy had small, but positive impacts on healthcare quality for commercially insured patients.

Chapter 2 examines the impact of a mandatory, multi-payer episode-based payment program on place-based racial disparities. The policy was implemented in 2013 in Medicaid and private insurance markets in Arkansas to reduce perinatal spending and improve quality for the "average" patient, but little is known about whether there were differential impacts on vulnerable populations, including racial minorities. Assessing place-based, rather than individual racial disparities, enables an assessment of how the interaction between race and socioeconomic drivers contributes to variation in policy effects across markets. Relative to FFS, episode-based payments shift financial risk to physicians by offering a single, risk-adjusted case rate for the entire course of care, rather than reimbursing providers separately for each service. Under this policy design, providers that keep costs below a risk-adjusted target threshold and meet certain benchmarks on quality measures earn a portion of the savings, while those that exceed it incur a financial penalty. Holding providers accountable for quality and costs of care during the episode creates a financial incentive to coordinate care and improve the patient experience across both dimensions.

In this chapter, commercial claims from the Truven MarketScan Database from 2010 to 2016 were used to examine whether the episode-based payment program differentially impacted physician behavior in areas with a high proportion of Black patients relative to areas with a high proportion of White patients. Difference-in-difference-in-differences analyses were conducted to assess how the policy affected quality of care across three phases of the care continuum, including prenatal, intrapartum, and postpartum services. Results show that in the short-term, there were significantly greater quality improvements in areas with a high proportion of White residents, but that areas with Black patients were able to close the gap in the long-term. This study aims to refine the understanding of the role that physician incentives play on racial disparities, which is a critically important question given current inequalities in healthcare and the potential for payment policies to help close or widen this gap.

Chapter 3 leverages a multi-payer episode-based payment program in Arkansas with compulsory provider participation for Medicaid and commercially insured enrollees, to test whether baseline quality was associated with heterogeneity in effects across quality and spending outcomes. In federal programs, bundled payment participation is typically voluntary, making it challenging to identify the impact of program participation on episode savings and quality improvement, particularly if providers with high quality at baseline differentially choose to participate. This policy is distinct from the conventional Medicare bundled payment structure by requiring participation across all providers and targeting younger patients insured by Medicaid and private payers. To date, little is known about how effects vary across providers with different baseline characteristics. It is important to explore variation in the impacts of such programs, since the "average" effect may attenuate towards the null if providers with different baseline characteristics experience divergent effects.

This chapter uses the Truven MarketScan commercial claims database from 2010 to 2016 in a difference-in-differences-in-differences design. I find that geographic areas with high performing providers at baseline were able to generate significantly greater savings through reduced hospital prices and volume of services. Geographic areas with low performing providers at baseline achieved larger improvement in quality measures directly linked to payment. However, neither of these differential effects were sustained in the long-term. The findings help to inform whether the effects of mandatory versus voluntary payment reform are expected to vary, to offer insight into future state and federal bundled payment design.

Together, these three papers form a body of work demonstrating the indirect impact of financial incentive design on physician behavior with respect to healthcare quality and spending among privately insured enrollees. Chapter 1 examines the spillover effects of a Medicaid payment reform that discontinued payment for a low-value service. Chapter 2 focuses on the role of a multi-payer bundled payment program on racial disparities, to gain a better understanding of how physician incentives can be used to address current inequalities in healthcare. Finally, Chapter 3 assesses whether episode-based payments with compulsory provider participation are associated with heterogeneity in effects by providers' baseline performance, across quality and spending outcomes.

These studies suggest that financial incentives remain an important determinant in shaping the amount and type of medical services that physicians supply, and need to be considered when designing physician payment reforms. This dissertation presents evidence on the importance of incentive design in perinatal care, particularly when considering the indirect impacts on specific patient populations, including beneficiaries insured by other payers and vulnerable populations. For racial minorities and patients that typically experience low quality of care, because of disproportionate barriers to accessing care and the potential for drastic improvement in health from treatment, financial incentives that facilitate quality improvement across the continuum of care are a necessity. The analyses in this dissertation can help to inform public policy to ensure that as value-based payment becomes more ubiquitous, physicians are incentivized to improve care for all populations.

1.4. References

Allen, L., & Grossman, D. (2019). The impact of voluntary and nonpayment policies in reducing early-term elective deliveries among privately insured and Medicaid enrollees. *Health Services Research*, 1–8. https://doi.org/10.1111/1475-6773.13214

Baicker, K., Chernew, M., & Robbins, J. (2013). *The Spillover Effects of Medicare Managed Care: Medicare Advantage and Hospital Utilization*. Retrieved from http://www.nber.org/papers/w19070.ack

Blumenthal, D., & Abrams, M. (2020). The Affordable Care Act at 10 Years — Payment and Delivery System Reforms. *New England Journal of Medicine*, *382*(11), 1057–1063. https://doi.org/10.1056/nejmhpr1916092

Carroll, A. E. (2017). Why Is US Maternal Mortality Rising? *JAMA*, *318*(4), 321. https://doi.org/10.1001/jama.2017.8390

Chee, T. T., Ryan, A. M., Wasfy, J. H., & Borden, W. B. (2016). Current state of value-based purchasing programs. *Circulation*, *133*(22), 2197–2205. https://doi.org/10.1161/CIRCULATIONAHA.115.010268

Chou, Y.-J., Huang, N., Lin, I.-F., Deng, C.-Y., Tsai, Y.-W., Chen, L.-S., & Lee, C.-H. (2006). Do Physicians and Their Relatives Have a Decreased Rate of Cesarean Section? A 4-Year Population-Based Study in Taiwan. *Birth*, *33*(3), 195–202. https://doi.org/10.1111/j.1523-536X.2006.00104.x

Consumer Reports. (2017). *Most U.S. Hospitals' C-Section Rates Exceeding National Targets*. Retrieved from https://www.consumerreports.org/media-room/press-releases/2017/05/consumer_reports_analysis_most_us_hospitals_csection_rates_exceeding_national_targets/

Dahlen, H. M., Mccullough, J. Mac, Fertig, A. R., Dowd, B. E., & Riley, W. J. (2017). Texas Medicaid Payment Reform: Fewer Early Elective Deliveries And Increased Gestational Age And Birthweight. *Health Affairs*, *36*(3), 460–467. https://doi.org/10.1377/hlthaff.2016.0910

Damberg, C. L., Elliott, M. N., & Ewing, B. A. (2015). Pay-For-Performance Schemes That Use Patient And Provider Categories Would Reduce Payment Disparities. *Health Affairs*, *34*(1), 134–142. https://doi.org/10.1377/hlthaff.2014.0386

Ellis, R. P., & McGuire, T. G. (1986). Provider behavior under prospective reimbursement. Cost sharing and supply. *Journal of Health Economics*, *5*(2), 129–151. https://doi.org/10.1016/0167-6296(86)90002-0

Fuchs, V. R. (1978). The supply of surgeons and the demand for operations. *The Journal of Human Resources*, *13 Suppl*, 35–56. https://doi.org/10.2307/145247

Gilman, M., Adams, E. K., Hockenberry, J. M., Milstein, A. S., Wilson, I. B., & Becker, E. R. (2015). Safety-net hospitals more likely than other hospitals to fare poorly under medicare's value-based purchasing. *Health Affairs*, *34*(3), 398–405. https://doi.org/10.1377/hlthaff.2014.1059

Greenberg, M. (2020). *Essential Data: America's Essential Hospitals 2018 Annual Member Characteristics Survey*.

Ikegami, N. (2015). Fee-for-service payment – an evil practice that must be stamped out? *International Journal of Health Policy and Management*, Vol. 4, pp. 57–59. https://doi.org/10.15171/ijhpm.2015.26

Joynt Maddox, K. E. (2018). Financial Incentives and Vulnerable Populations — Will Alternative Payment Models Help or Hurt? *New England Journal of Medicine*, *378*(11), 977–979. https://doi.org/10.1056/NEJMp1715455

Kamal, R., & Cox, C. (2018). *How has U.S. spending on healthcare changed over time?* Retrieved from https://www.healthsystemtracker.org/chart-collection/u-s-spending-healthcare-changed-time/?_sf_s=changed

Kessel, R. A. (1958). Price Discrimination in Medicine. *The Journal of Law and Economics*, *1*, 20–53. https://doi.org/10.1086/466540

Khullar, D., Song, Z., & Chokshi, D. A. (2018). Safety-Net Health Systems At Risk: Who Bears The Burden Of Uncompensated Care? *Health Affairs Blog*. Retrieved from https://www.healthaffairs.org/do/10.1377/hblog20180503.138516/full/

Kozhimannil, K. B., Hardeman, R. R., & Henning-Smith, C. (2017). Maternity care access, quality, and outcomes: A systems-level perspective on research, clinical, and policy needs. *Seminars in Perinatology*. https://doi.org/10.1053/j.semperi.2017.07.005

MacDorman, M. F., Declercq, E., Cabral, H., & Morton, C. (2016). Recent Increases in the U.S. Maternal Mortality Rate: Disentangling Trends From Measurement Issues. *Obstetrics and Gynecology*, *128*(3), 447–455. https://doi.org/10.1097/AOG.0000000000001556

Mcguire, T. G. (2000). Physician Agency. In *Handbook of Health Economics* (Volume 1, Part A, pp. 461–536). https://doi.org/10.1016/S1574-0064(00)80168-7

MedPAC. (2019). *A data book: health care spending and the Medicare program*. Retrieved from www.medpac.gov.

MedPAC. (2020). *Assessing payment adequacy and updating payments in fee-for-service Medicare*.

Mendelson, A., Kondo, K., Damberg, C., Low, A., Motúapuaka, M., Freeman, M., … Kansagara, D. (2017). The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care. *Annals of Internal Medicine*, *166*(5), 341. https://doi.org/10.7326/M16-1881

Rama, A. (2017). *Policy Research Perspectives Payment and Delivery in 2016: The Prevalence of Medical Homes, Accountable Care Organizations, and Payment Methods Reported by Physicians*. Retrieved from https://fred.stlouisfed.org/series/POPTHM.

Rice, T. H. (1983). The impact of changing medicare reimbursement rates on physician-induced demand. *Medical Care*, *21*(8), 803–815. https://doi.org/10.1097/00005650-198308000-00004

Sachs, R. M., Burn, S., Conti, R., Dafny, L., Dague, L., Devlin, A., … Ristovska, L. (2019). *Safety Net Cutbacks and Hospital Service Provision: Evidence from Psychiatric Care*.

Town, R., Wholey, D. R., Kralewski, J., & Dowd, B. (2004). Assessing the Influence of Incentives on Physicians and Medical Groups. *Medical Care Research and Review*, *61*(3), 80–118. https://doi.org/10.1177/1077558704267507

Truven Health Analytics. (2013). *The Cost of Having a Baby in the United States*. Retrieved from http://www.chqpr.org/downloads/CostofHavingaBaby.pdf

2. Understanding Spillover Effects of Medicaid Payment Reform for a Low-Value Service in the Privately Insured

2.1. Introduction

Physician behavior is a key driver of health outcomes and spending, as many medical services are not directly demanded by a patient, but requested by a doctor on the patient's behalf (Chandra & Skinner, 2012; Ellis & McGuire, 1986; Kessler & McClellan, 1996; Smith, Saunders, Stuckhardt, & McGinnis, 2013). One factor driving physician decision-making is the financial incentive that he or she receives. It is well established that under the prevailing payment model, fee-for-service (FFS), physicians are influenced to overprescribe low-value care to maximize their income (Ellis & McGuire, 1986; Ikegami, 2015; Mendelson et al., 2017). Low-value care, defined as the provision of a medical procedure that provides little or no benefit to patients, deviates from the social optimum, and has potential to cause harm or incur unnecessary costs (Brownlee et al., 2017; Brownlee, Saini, & Cassel, 2014; Mafi et al., 2017; Maratt et al., 2019; McGlynn et al., 2003). To address this concern, many states and payers have adopted policies that move payment systems away from traditional FFS towards reimbursement that aligns spending with quality. These approaches often aim to discourage the provision of low-value care by incorporating lower physician earnings for these services (MedPAC, 2019). While a large body of literature examines physicians' direct response to these payment incentives, few papers focus on indirect responses such as whether there are spillover effects to payers and services outside of the target population.

Economic theory suggests that physicians may respond to lower payment for a low-value service by reducing the provision of the service across all payers (e.g. a positive spillover) and/or by increasing the provision of a more profitable service to compensate for lost income (e.g. inducing demand), but there is little empirical evidence on this question. Understanding these indirect effects is critical to developing a full understanding of physician behavior under reimbursement changes. Positive spillovers indicate that the desired physician behavior can have welfare-improving benefits beyond the population of interest, while physician-induced demand is an unintended consequence that can introduce harm to patients. Whether these indirect behavioral responses transpire concurrently, and the drivers of these effects, can help policymakers to determine the far-reaching implications of payment reform, and the contexts in which physicians are likely to customize or standardize care across patients.

This paper explores how discontinuing Medicaid payment for Early Elective Deliveries (EEDs), a low-value mode of childbirth defined as a scheduled, non-medically necessary induction or cesarean section (c-section) prior to 39 weeks gestation, affects privately insured patients. I use the Hospital Compare and Truven MarketScan Commercial Claims databases to address the following questions, comparing the effect of Medicaid nonpayment to other financial and non-financial incentives in a difference-in-differences (DD) analysis: (1) does the change in Medicaid payment prompt a payer-related spillover, in the form of a reduction in EEDs in the commercial sector? (2) is there evidence of physician-induced demand through an increase in low-risk c-sections, which are less time-intensive and more profitable, but riskier than vaginal deliveries? and (3) are variations in spillovers consistent with financial or reputational drivers?

In perinatal care, fee reductions associated with low-value EEDs emerged as a promising option to reduce long-term hospital costs. In 2010, EEDs constituted nearly 20% of all U.S. hospital births, exceeding the patient safety target of 5% (Main et al., 2010). Because EEDs can be conveniently scheduled and are paid at the same rate as full-term deliveries under FFS, there are strong incentives to continue their provision. EEDs pose significant dangers to mothers, including increased risk of infection and postpartum hemorrhage. They have no known clinical benefits, but may be appealing to physicians to enable a convenient schedule and to mitigate perceived liability concerns (Choosing Wisely, 2013). However, EEDs often generate higher medical expenses compared to full-term, spontaneous births.

Since Medicaid covers 45% of births, Medicaid programs in Georgia, Indiana, Missouri, and Mississippi adopted policies that discontinued physician reimbursement for EEDs between January 1, 2014 and January 1, 2015 (The Henry J. Kaiser Family Foundation, 2019). Physicians were only eligible to receive payment for births prior to 39 weeks' gestation if it was properly documented as medically necessary. This provided a financial incentive to reduce overuse of this service because continuing its supply would forego the average $7,213 that a physician earns per Medicaid birth (Caughey et al., 2009). Several states adopted a range of other approaches. Some states implemented Medicaid pay-for-performance programs, which provided physicians with a bonus payment if they achieved a goal EED rate. Other states implemented non-financial approaches, including voluntary "hard stop" initiatives, which encouraged hospitals to take a pledge to end the provision of EEDs, and quality improvement collaboratives, which aimed to educate physicians and expecting mothers about the dangers of EEDs through multi-stakeholder advocacy and public reporting.

I investigate indirect effects of the Medicaid payment policy on low-value care among the commercially insured, and compare effects to four groups with varying financial and non-financial incentives, including: (1) no policy aimed at curbing EEDs, (2) a voluntary "hard stop" initiative, (3) a quality improvement collaborative, and (4) a Medicaid pay-for-performance program. To date, studies have measured the Medicaid policy's direct effects (Allen & Grossman, 2019; Dahlen et al., 2017). These studies find a mean 16% decline in Medicaid EEDs. Prior work has not examined indirect effects of the policy in the non-Medicaid population, nor has it compared the impact of the Medicaid payment policy to other financial and non-financial approaches. I aim to fill this gap by exploring whether effects of the Medicaid nonpayment policy spilled over to the commercial sector, and how these effects varied by policy approach.

I find that the observed change in Medicaid EED volume extended to the commercial sector. All-payer EEDs declined 3.35% in treatment states relative to states with no policy to reduce EEDs. The Medicaid payment policy also lowered all-payer EEDs by 3.92% and 3.58% compared to states with a hard stop policy and a pay-for-performance payment program, respectively. Hospital Compare data do not allow me to disentangle the direct result in Medicaid from a spillover to privately insured patients; to gain insight into this question, I examine whether effects increased in geographic areas with a higher share of Medicaid patients. Effects did not vary with the Medicaid rate, providing suggestive evidence that commercial patients were impacted.

I also find no evidence of inducement-related spillovers, under which the income lost from EEDs would prompt an increase in other low-value services (e.g. low-risk c-sections) (McGuire & Pauly, 1991). On average, reimbursement for c-sections is 50% higher than vaginal deliveries. C-sections can also be dangerous for low-risk women (Teleki, 2017). Given that there were no significant changes in low-risk c-sections, there is no evidence that physicians compensate for lost income by increasing the provision of more profitable services.

Finally, I find a larger positive spillover in areas with a higher share of for-profit hospitals. Physicians in for-profit and non-profit hospitals might have different priorities (e.g. practitioners in non-profit hospitals may have reputational objectives, while those in for-profits have financial ones), suggesting that the response to financial incentives may vary (Dranove, Garthwaite, & Ody, 2017; Horwitz, 2005; Newhouse, 1970). This result may indicate a stronger response to financial incentives by physicians in a for-profit setting, where physician and hospital financial objectives are aligned. It may also highlight barriers to changing physician practice patterns when their objectives are reputational, rather than financial.

This study makes several contributions to the literature. First, this analysis is one of the first to study payer- and service-related spillovers of financial incentives in the same setting. This examination lends insight into how physicians customize care, an important factor in understanding the full effect of reimbursement changes. Second, Medicaid nonpayment provides a unique opportunity to study a major incentive to reduce low-value care. Most value-based payments cap downside risk at approximately 3%, while the policy of interest imposes a 100% penalty. Third, this policy presents a novel opportunity to study indirect effects of financial incentives on physician behavior, and to understand the impact at a system level. Studying a physician's financial incentives in the context of perinatal care has several advantages. Perinatal care is clinically salient. Since 2013, average inpatient costs have increased 32%, while overall hospital spending has only grown by 4.8% (Kamal & Cox, 2018; Truven Health Analytics, 2013). Nonetheless, the U.S. experiences the highest maternal mortality rate of all developed countries (Carroll, 2017). Over half of U.S. hospitals lag below the national target in quality scores, suggesting much room for improvement (Consumer Reports, 2017). Volume is also relatively independent from physician influence (e.g. compared elective arthroplasty, which requires a pre-surgical appointment), and repeat birth is unpredictable. Thus, the physician response is expected to stem directly from the payment reform as opposed to potential confounders (Carroll, Chernew, Fendrick, Thompson, & Rose, 2018), increasing the potential to generalize results to other settings. Perinatal care has features linked to strong provider responses, including variation in quality and costs, and presence of elective services with practical treatment substitutes (Chou et al., 2006).

I show that reducing Medicaid payment for low-value services provides two broad results: (1) it discourages overprovision of low-value care in the non-Medicaid population, and (2) it leads to modest spillovers to the non-Medicaid population, which are smaller than the direct effect in Medicaid. I further illustrate that physicians are generally more responsive to financial incentives when it aligns with hospital objectives. I also highlight important tools for incentive design, including: (1) a financial penalty for low-value care can lead to stronger spillovers than a financial bonus, (2) a mandatory incentive can be more effective than a voluntary one, and (3) a policy with interdisciplinary collaboration and educational components can be as successful as a financial incentive. These comparisons provide greater understanding of payment reform, such as how to structure incentives, determine participation, and incorporate education and teamwork elements.

The paper proceeds as follows. Section 2.2 provides background on incentives including the policy landscape and prior literature. Section 2.3 lays out the theoretical framework. Section 2.4 describes the analytic dataset and empirical strategy. Section 2.5 discusses the main results. Section 2.6 concludes.

2.2. Background

2.2.1. Literature on Indirect Physician Responses to Financial Incentives

There is a large body of economic literature investigating how physicians, and healthcare providers more generally, respond to financial incentives. Two distinct streams of literature focus on indirect effects; this includes spillovers, or the extent to which a financial incentive directed towards patients with one insurance type affects patients with other insurance types (Frank et al., 2007; Glied & Zivin, 2002; Tai-Seale, McGuire, & Zhang, 2007), and physician-induced demand, where physicians, acting as agents on behalf of patients (who lack medical knowledge to make autonomous treatment decisions), request additional volume of services in response to negative income shocks (Evans, 1974; McGuire & Pauly, 1991).

First, spillovers are difficult to predict, as there is debate over whether physicians adhere to “custom made” or “ready-to-wear” treatments (e.g. addressing an individual patient’s needs on a case-by-case basis or treating a broad class of patients with a standardized “norm,” respectively) (Frank et al., 2007). On one hand, doctors have demonstrated an inclination to treat a “modal” patient, rather than differentiate by insurer, to circumvent communication, cognition, coordination, and capability costs. This suggests that when a physician’s dominant payer alters payment incentives, the behavioral response may spillover to other populations (Frank et al., 2007; Glied & Zivin, 2002; Tai-Seale et al., 2007). Spillovers from Medicare to privately insured patients have also been observed with non-financial incentives (Barnett, Olenski, & Sacarny, 2020). On the other hand, empirical work supports the notion that physicians customize care across patients, as varying payment rates from public and private insurers have led to significantly different utilization patterns, waiting times, and number of follow-up visits (Jürges, 2009; Lungen, Stollenwerk, Messner, Lauterbach, & Gerber, 2008; Newhouse & Marquis, 1978; Schwierz, Wübker, Wübker, & Kuchinke, 2011). Despite evidence in both areas, there is little understanding of the circumstances under which physicians use custom made versus ready-to-wear treatments.

Developing a greater insight into the factors that lead to variation in spillovers is critical to understanding physician responses to payment reforms.

Second, prior studies on physician-induced demand support this theory for Medicare and commercial payment changes, particularly when procedures are intense and elective (e.g., heart attack treatment, advanced imaging, and c-sections) (Clemens & Gottlieb, 2014; Coey, 2015; Foo, Lee, & Fong, 2017; Gruber & Owings, 1996; Jacobson, Earle, Price, & Newhouse, 2010; Yip, 1998). Under these conditions, inducement becomes more desirable because the profit margin tends to be high and the service tends to require minimal time. In contrast, the limited literature on fee changes in Medicaid have found no evidence of inducement, arguing that non-Medicaid findings cannot be generalized since Medicaid patients are a small share of a physician's patient pool and prices tend to be low (Alexander, 2015; Gruber, Kim, & Mayzlin, 1999). More research on inducement and other unintended consequences of Medicaid incentives is imperative for developing state policies aimed at reducing low-value care. I expand the empirical literature on spillovers and induced demand by examining concurrent dynamics between substitution and income effects from Medicaid to commercial patients.

The magnitude of indirect effects may vary based on financial and/or reputational market factors. For example, a physician's objectives may change depending on whether they are employed by a for-profit or non-profit hospital. Theoretically, non-profits seek to maximize quality and income, while for-profits are predominantly focused on income (Newhouse, 1970). Thus, for-profit hospitals tend to be more sensitive to payment changes. Prior studies have demonstrated that for-profit hospitals are more likely to offer high-margin services than non-profit hospitals. For-profit hospitals also varied the quantity of a given service with price changes over time, suggesting greater responsiveness to service profitability (Dranove et al., 2017; Horwitz, 2005). These differences intensified with hospital size and market competition (Horwitz, 2005; Moon & Shugan, 2020; Rosko, Al-Amin, & Tavakoli, 2020). However, because non-profit hospitals offer a wider range of services (both profitable and unprofitable), they may have a competitive advantage over for-profit hospitals. As a result, overall differences in profits by hospital type are usually insignificant (Lakdawalla & Philipson, 1998, 2006; Moon & Shugan, 2020). There is little research on how physicians employed by different hospital types respond to value-based payment changes, especially with respect to indirect effects.

2.2.2. Policy Landscape on Incentives in Perinatal Care

At the federal level, providers are only encouraged by non-financial means to reduce EEDs. In February 2013, the Choosing Wisely campaign, in conjunction with the American College of Obstetricians and Gynecologists (ACOG), released an official federal guideline discouraging EEDs. Since 2007, several states adopted policies employing financial or non-financial incentives aimed at reducing EEDs across all payers, in an effort to generate perinatal care cost savings and improve birth outcomes.

In this study, I evaluate the effect of a state level Medicaid policy change that stopped physician payment for EEDs. The first state to implement this strategy was Texas in 2011. Since, ten other states have enacted the same policy (New York, New Mexico, Nevada, Montana, South Carolina, Louisiana, plus the four treatment states). Reducing or eliminating payments for a given service is advantageous from a policy perspective due to its simplicity. It is methodologically straightforward, which serves as a strong predictor of effectiveness. Studies show that incentives are most conducive to behavior change and decreased gaming when there is a clear, one-to-one relationship between the behavior and reinforcement (Town et al., 2004). Further, this payment policy largely retains a FFS structure, the preferred payment by physicians (Bain, 2017; Ikegami, 2015). I compare the spillover effects of this policy to three other policies that have similar goals and offer either financial or non-financial incentives to physicians.

Another approach that leverages financial incentives is Medicaid pay-for-performance reimbursement, which offers a bonus payment to physicians that achieve a benchmark EED rate. In 2010, Washington launched the Safety Net Assessment Act, which gave hospitals a 1% increase in their Medicaid reimbursement for reducing EEDs from one year to the next (Association of State and Territorial Health Officials, 2014). Colorado adopted a similar program in 2011, called the Hospital Quality Incentive Payment (HQIP) Program. HQIP offers volume-adjusted payments based on Medicaid discharges and quality achievement on five performance measures (one of which is EEDs) (Colorado Medicaid, 2016). Wisconsin rolled out the Obstetric Medical Home (OBMH) program between 2011 and 2013, which pays the obstetrician an additional $1,000 for each Medicaid patient that attends ten prenatal visits and a postpartum visit within 60 days of birth. OBMH practitioners are given an additional $1,000 bonus per positive birth outcome, including full-term births (Agrawal, 2017). Differences between discontinuing payment for EEDs and providing a bonus for EED performance in Medicaid applies behavioral economic principles of prospect theory, under which individuals value gains and losses of the same magnitude asymmetrically. In other words, they lose more utility from a penalty than they gain from an equivalent bonus. From a behavioral perspective, this suggests that when faced with uncertainty (e.g. reimbursement changes for EEDs), financial penalties will be more effective than bonuses; physicians, aiming to avoid financial losses, will reduce the provision of low-value care, even if it means suboptimal expected utility (Kahneman & Tversky, 1979).

There are two main approaches that utilize non-financial incentives. The first is a "hard stop" policy, under which hospitals voluntarily pledge to ban EEDs by requiring hospital review and approval for any delivery before 39 weeks' gestation without documented indication. Eleven states (Arkansas, Arizona, California, Delaware, Iowa, Michigan, Minnesota, North Carolina, Oregon, Tennessee, and Oklahoma) implemented these initiatives between 2009 and 2013. This policy adds an effort-related cost to providing an EED to discourage their use. These programs target hospitals with a relatively higher share of Medicaid births, but all hospitals are encouraged to participate. The pledge to reduce EEDs applies to the overall hospital rate, and is not payer-specific. The second approach is a quality improvement collaborative, which takes on a range of structures. They generally involve a coalition of professional, clinical, and non-governmental organizations rolling out educational awareness campaigns on the dangers of EEDs and low-risk c-sections for expecting mothers and delivering obstetricians. Many programs go a step further by requiring hospitals to report their EED rates, to track performance and promote accountability. This strategy is two-pronged, as it attempts to change the culture surrounding provision of EEDs through education and multi-stakeholder buy-in, while also publicly comparing physicians with their peers to enact social pressure and change norms.

2.2.3. Effects of the Medicaid Payment Policy

Empirical studies have measured the Medicaid policy's direct effects in single state analyses using a DD framework (Allen & Grossman, 2019; Dahlen et al., 2017). Dahlen et al. (2017) examined the impact of the 2011 Texas Medicaid payment change implemented concurrently with a voluntary hard stop policy, finding a 14% significant decline in the EED rate compared to control states. The largest impacts were observed among minority patients. Birth outcomes also improved significantly, with birthweight increasing by 6 ounces (Dahlen et al., 2017). Allen and Grossman (2019) explored the effects of the Medicaid policy and a voluntary hard stop policy in South Carolina. The study found that the payment policy (which was implemented in both commercial and Medicaid markets) reduced EEDs by 16.6%, while the hard stop policy reduced them by 12.7%, relative to controls. The decline in EEDs was higher for Medicaid than commercial patients (18.9% versus 16.6%, respectively) (Allen & Grossman, 2019). Byanova (2015) also assessed the joint effects of a hard stop initiative and Medicaid nonpayment in Texas, finding that Medicaid and non-Medicaid EEDs declined by 18.5% and 5.9%, respectively. The study also observed a 13.1% increase in the non-Medicaid total c-section rate, attributing this to demand inducement (Byanova, 2015). Existing work has not examined indirect effects of the policy in the non-Medicaid population, and no studies have adopted a multi-state analysis. Further, studies have not directly compared the impact of the Medicaid payment policy to other financial and non-financial approaches, or examined heterogeneity across financial and reputational market characteristics.

2.3. Theoretical Framework

2.3.1. Model

In this section, I develop a theoretical framework for understanding why variation in spillovers occurs in response to financial incentives, in the context of low-value services. In particular, I am interested in how physicians adjust the provision of care in the commercial sector in response to Medicaid nonpayment of a low-value service. I begin with a physician utility model in the style of Ellis and McGuire (1986), where a physician selects a quantity of services to maximize utility over profits (π) and patient well-being (B) (Ellis & McGuire, 1986). I extend the framework by considering how a physician may vary behavior across payers and services. In particular, the utility model U allows flexibility for a physician to choose a different quantity of care for Medicaid (x_m) and non-Medicaid (x_n) patients across services $j = 1,2, \dots, k$:

$$U = \sum_{j=1}^{K} \alpha_{\text{HOSP}} \left[B\left(x_{j,m}\right) + B\left(x_{j,n}\right) \right] + \delta\pi\left(x_{j,m}\right) + (1-\delta)\pi(x_{j,n}) \qquad [1]$$

In [1], profit is a function of the income for Medicaid and non-Medicaid services (ρ_m and ρ_n), defined by net earnings after subtracting monetary costs from total reimbursement; non-financial, or implicit, effort-related costs (e); and quantity of care selected across Medicaid and non-Medicaid patients (x_m and x_n):

$$\pi(\rho, e; x_m, x_n) \qquad [2]$$

The model adds two unique components. First, it considers the share of a physician's patients that are insured by Medicaid, δ, and the remaining proportion insured by non-Medicaid payers $(1 - \delta)$. Second, it assumes that the weight a physician places on patient benefits varies by the type of hospital that a physician is employed. In practice, a physician's preference contains a weight for agency, α, which represents the marginal rate of substitution, or the rate at which the physician is willing to trade off one dollar of hospital profit for one dollar of patient benefit, such that $1 > \alpha > 0$. If a physician were to serve as perfect agent for the patient, then $\alpha = 1$, and then the physician weights their profit equally to the patient's benefit. I assume that α varies with the index HOSP, or the type of hospital that a physician is employed (non-profit, for-profit, or public), and that all else equal, the physician serves as a better agent when employed by a non-profit, relative to a for-profit, hospital ($\alpha_{\text{FORPROF}} < \alpha_{\text{NONPROF}}$).

For simplicity, I explore the case where *k* represents *E* for EEDs, *C* for low-risk c-sections, and *V* for full-term vaginal deliveries. The distinction between these services is that *E* and *C* are low-value, and *V* is high-value, so the implicit costs of providing *E* and *C* (e.g. concerns about harming the patient, uncertainty as to whether the service is appropriate) are relatively higher. The marginal profit is higher for *C* than for the other services. The first-order conditions (FOCs) for Medicaid and non-Medicaid treatment are:

$$\alpha_{\text{HOSP}}\left[B'\left(x_{E,m}\right) + B'\left(x_{C,m}\right) + B'\left(x_{V,m}\right)\right] + \delta\left[\pi'\left(x_{E,m}\right) + \pi'\left(x_{C,m}\right) + \pi'\left(x_{V,m}\right)\right] = 0 \quad [3]$$

$$\alpha_{\text{HOSP}}\left[B'\left(x_{E,n}\right) + B'\left(x_{C,n}\right) + B'\left(x_{V,n}\right)\right] + (1-\delta)\left[\pi'\left(x_{E,n}\right) + \pi'\left(x_{C,n}\right) + \pi'\left(x_{V,n}\right)\right] = \quad [4]$$

In general, I assume that $\pi'\left(x_{j,n}\right) > \pi'\left(x_{j,m}\right)$, or the marginal profit is higher for non-Medicaid patients for a given service, as prices tend to be higher on average. Assuming that marginal patient benefits are equal across payers (e.g. $B'\left(x_{E,m}\right) = B'\left(x_{E,n}\right)$, $B'\left(x_{C,m}\right) = B'\left(x_{C,n}\right)$, and $B'\left(x_{V,m}\right) = B'\left(x_{V,n}\right)$), and rearranging:

$$\frac{\delta}{(1-\delta)} = \frac{\pi'(x_{E,n})+\pi'(x_{C,n})+\pi'(x_{V,n})}{\pi'(x_{E,m})+\pi'(x_{C,m})+\pi'(x_{V,m})} \quad [5]$$

When the Medicaid payment policy is implemented, the marginal profit for Medicaid EEDs, $\pi'\left(x_{E,m}\right)$, drops significantly. Since the payment policy is only implemented in Medicaid, the marginal profit for non-Medicaid EEDs remains relatively higher, and the gap widens relative to pre-policy implementation: $\pi'\left(x_{E,m}\right) < \pi'\left(x_{E,n}\right)$. Intuitively, the volume of Medicaid EEDs will fall with the price, a result that has been demonstrated empirically (Allen & Grossman, 2019; Byanova, 2015; Dahlen et al., 2017). Since the right and left terms are equivalent, it follows that the total marginal profit among non-Medicaid services, $\pi'\left(x_{E,n}\right) + \pi'\left(x_{C,n}\right) + \pi'\left(x_{V,n}\right)$, increases with δ, or share of Medicaid patients. Based on this result, it is likely that the mix of services changes among the non-Medicaid population after implementation of the Medicaid payment policy, but whether and how physicians choose to substitute depends on the spillover mechanism.

2.3.2. Mechanisms for Spillovers

Spillovers may arise through several mechanisms. One likely avenue is through use of ready-to-wear treatments, or common practice patterns across patients, regardless of insurance type (Frank et al., 2007). This may transpire as a physician's strategy to combat the challenges of customizing care for patients insured by different payers. It may also result from learning new skills while treating Medicaid patients, and applying them to non-Medicaid patients (Baicker et al., 2013; Chernew, Baicker, & Martin, 2010). In the model, this is represented by marginal profits for EEDs, $\pi'(x_{E,n})$, decreasing in δ because of implicit, non-monetary costs of continuing to provide the service. The magnitude of the response is expected to rise as a physician gains Medicaid patients, with practice patterns converging towards the "modal" patient (Glied & Zivin, 2002). Spillovers are also more likely to arise when patients have similar clinical reasons for seeking care (e.g. childbirth). In these circumstances, it is likely that the physician can use the same standards of care across patients, potentially leading to improvements in outcomes and cost savings (Chernew et al., 2010). If a reduction in non-Medicaid EEDs is accompanied by an increase in full-term vaginal deliveries, the spillover may be a welfare-improving, as it offsets services where the marginal cost exceeds net patient benefit (Baicker et al., 2013).

Another channel for spillovers is through physician-induced demand, under which physicians respond to negative income shocks by increasing volume or intensity of services, beyond the optimal amount (McGuire & Pauly, 1991). Inducement is "costly" for the physician, in the sense that it may cause harm to the patient; thus, it will only occur when the profit margin is high and the time cost is low. Spillovers of a fee change in one market may be characterized by inducement in more profitable sectors, since the physician has alternate avenues through which to recoup income. In particular, inducement is likely to occur when a substitutable, more profitable, and less time intensive service exists through which to pursue inducement (Chernew et al., 2010; McGuire & Pauly, 1991). In the model, I may observe physician-induced demand through the effect of the Medicaid payment policy on low-risk c-sections among non-Medicaid patients. In this setting, it is possible that the physician seeks to recover income lost for EEDs by increasing the volume of a more profitable service, such as low-risk c-sections. As the marginal profit for low-risk c-sections increases compared to its substitutes, inducement-related spillovers are expected to rise. These effects are also likely to amplify when exposure to the fee change, measured by δ, increases, because physicians must induce more to account for a greater proportion of lost income. This spillover suggests that policy efforts to reduce spending in Medicaid may be cost-increasing in the commercial sector, due to higher profitability among privately insured patients.

2.3.3. Physician Behavior by Hospital Type

The degree to which spillovers occur may be influenced by the physician's hospital type. Non-profit hospitals have a distinct objective function, under which they aim to maximize profits and quality, rather than focusing predominantly on profits (Newhouse, 1970). This is represented in the model by varying physician agency, α, by hospital type, where patient benefits are considered more important by physicians in a non-profit hospital setting ($\alpha_{\text{FORPROF}} < \alpha_{\text{NONPROF}}$). In general, this framework suggests that in the absence of financial incentives, physicians in non-profit hospitals maintain higher levels of quality, and lower supply of low-value services, relative to those in for-profits. It also implies that practitioners in non-profit hospitals may be less attentive to changes in the financial environment, including payment incentives aimed at reducing low-value care. In contrast, physicians in for-profit hospitals, driven by financial objectives, are likely to respond more strongly to changes in service profitability (Dranove et al., 2017; Horwitz, 2005).

2.3.4. Testable Predictions

In this analysis, I empirically test model predictions by investigating whether there were spillovers of EEDs and low-risk c-sections to the commercial sector after implementation of the Medicaid payment policy.

2.3.4.1. First, I test whether spillovers are prompted by ready-to-wear treatments.

Under this mechanism, I expect there to be a greater reduction in EEDs as policy exposure, or the share of Medicaid patients, rises. This would signal convergence of physician practice patterns towards a modal patient.

2.3.4.2. Second, I test if spillovers arise from physician-induced demand.

In this case, I expect the Medicaid payment policy to increase low-risk c-sections among the commercial population, as this service is less time consuming and more profitable compared to EEDs. Further, I anticipate the effects of induced demand to be higher in areas where c-sections are more profitable relative to vaginal deliveries, and where the magnitude of profit loss is greatest (e.g. in areas with more Medicaid patients).

2.3.4.3. Third, I test if variation in spillovers is driven by a physician's hospital setting.

If so, I expect to observe a stronger reduction in EEDs in areas with a greater share of for-profit, rather than non-profit and public hospitals, where objectives are financial, and physicians are more likely to respond to changes in service profitability.

2.4. Methods

2.4.1. Analytic Data Set

The data for the main outcome, EED rate, is Medicare's Hospital Compare Database from 2013 to 2017. Hospital Compare is a comprehensive, hospital-level database containing quality measures reported as part of a mandatory initiative, the Medicare Inpatient Quality Reporting (IQR) Program. It was formed through a public-private collaboration between Medicare and the Hospital Quality Alliance in 2002, to improve quality of U.S. hospitals by making information on hospital performance publicly accessible to consumers. Hospital Compare data are collected and updated from hospitals on a quarterly basis; all hospitals are required to submit rates for a set of quality measures, except critical access hospitals, which may voluntarily submit their data.

Hospital Compare data confer several advantages. First, all measures are validated and endorsed by the National Quality Forum (NQF), the only consensus-based healthcare organization in the U.S., as defined by the Office of Management and Budget. NQF endorsement is the gold standard for quality metrics, as it uses a transparent, evidence-based, and consensus-based process driven by experts. Second, the mandatory nature of the Hospital Compare data allow longitudinal, hospital-level data to be easily accessible for research. Unlike voluntary reporting, this ensures that there are no systematic omissions in the data set, which improves representativeness and generalizability. Finally, Hospital Compare relies on multiple data elements, including administrative data and medical records, which can improve completeness and accuracy. The data also have some limitations. All measures are aggregated to a single all-payer rate, so insurer- and physician-specific rates cannot be identified. Further, since hospitals are responsible for self-reporting data, differences in hospital size, types of patients, and sampling strategies may reduce standardization, and thus, measure precision may be limited. Finally, Hospital Compare began publicly reporting perinatal care measures in 2013, which limits the length of the pre-policy period (Centers for Medicare and Medicaid Services, 2016; National Quality Forum, 2019).

To create the analytic data set, I merge the Hospital Compare database to several other data sources. All data sets are aggregated to the Metropolitan Statistical Area (MSA) level and linked using a concatenated State-MSA identifier to obtain a consistent unit of analysis. State-MSA is the smallest common indicator across all data sets. To aggregate the Hospital Compare data, I use the Department of Housing and Urban Development's U.S. Postal Service's Zip Code Crosswalk. Since Zip Codes often overlap the boundaries of multiple MSAs, duplicate Zip Code records exist in the crosswalk (one per Zip Code-MSA pair). To ensure that the EED rate is proportionately counted within and across MSAs, and to account for population differences within a Zip Code, I multiply each EED numerator and denominator by the ratio of residential addresses as a weight. A residential ratio weight is available for each Zip Code-MSA pair, with the sum of weights for each Zip Code equaling 1.0 (Wilson & Din, 2018).

For the second outcome variable, low-risk c-sections, I use the Truven Marketscan Commercial Claims Database, which links paid claims and encounter data with detailed patient information across sites and types of providers over time. Data are available for an extended pre-policy period, so I use data from 2010 to 2017. The MarketScan database is a convenience sample of enrollees in commercial employer and health plans, but it includes proprietary claims from over 36 million hospital discharges (Johns Hopkins, 2016). This enables identification of the change in low-risk c-sections among the privately insured. These data are collected across broad geographic areas to represent treatment patterns and costs in the U.S. I use several maternal and clinical characteristics as covariates. One limitation is that a major insurer dropped out of the MarketScan data in 2015. To avoid differential selection into the database over time, I limit the sample to the employer population, which remains stable over the study period.

For the remaining covariates, I link to additional data sources (the U.S. Census Bureau's American Community Survey (ACS), the Health Resources and Services Administration's Area Health Resource File (AHRF), the American Hospital Association (AHA) Annual Survey, and the National Practitioners Data Bank (NPDB)). The ACS has publicly available data on demographic and employment characteristics for all counties. AHRF includes data on health professions and facilities, hospital utilization, and spending. Both are extracted at the county-level, and then aggregated and linked at the State-MSA level. The AHA data contains hospital-level information, so I collapse it to the State-MSA level using a county to MSA crosswalk, and then link it to the main data set. The AHA Survey is the most widely used database for hospital-level information, with an average response rate of 83%. The AHA estimates certain measures for non-reporting hospitals, or those that submit incomplete survey information, using U.S. Census and other national-level data sources. Finally, I use the state-level NPDB, a web-based repository containing information on state-level medical malpractice payments (AHA, 2018; HRSA, 2019; NPDB, 2019; U.S. Census Bureau, 2018).

The final analytic sample consists of 3,951 State-MSA-quarters between 2013 and 2017. This includes 553 State-MSA-quarters among treatment states; 436 State-MSA-quarters among states with no EED-related policies; 1,135 State-MSA-quarters among states with hard stop policies; 1,352 State-MSA-quarters among states with quality improvement coalitions; and 475 State-MSA-quarters among states with Medicaid pay-for-performance programs.

2.4.2. Empirical Strategy

The identification strategy is a difference-in-differences (DD) framework comparing treatment states (GA, IN, MO, and MS) to four sets of control states, including those with: (1) no policy to curb EEDs, (2) a voluntary hard stop policy, (3) a quality improvement collaborative, and (4) a Medicaid pay-for-performance program. The analysis spans one pre-implementation year (2013) and three post-implementation years (2015-2017). The EED Medicaid policy acts as a source of exogenous variation, where physicians in GA, IN, MO, and MS are subject to discontinued reimbursement for Medicaid EEDs, while control states are not, leading to a quasi-experimental design. This approach is modeled on earlier work comparing direct effects of the Medicaid payment policy on utilization and birth outcomes within a single state (Allen & Grossman, 2019; Byanova, 2015; Dahlen et al., 2017). Each state implemented the Medicaid payment policy between January 1, 2014 and January 1, 2015. Due to the variation in treatment timing, I begin the post period in 2015. By omitting 2014 (the period in which states were launching the Medicaid payment policy at different time points), I ensure that each state has the same pre- and post-period. This aims to avoid biasing estimates by including later-treated states in the comparison group before treatment begins (Goodman-Bacon, 2018).

The control groups represent types of initiatives aimed at curbing EEDs in other states. This strategy is twofold. First, it mitigates the threat of contaminating the true effect of the Medicaid payment policy of interest by directly comparing states that have employed a single approach. It also enables comparing the Medicaid payment policy to other financial and non-financial incentive structures. The policies in each comparison group were enacted prior to 2014. The main control group includes eight states with no policies in place to reduce EEDs (ID, ME, NE, NJ, RI, ND, SD, VA, and WY). The Medicaid pay-for-performance comparison group has three states (WA, CO, and WI). For non-financial incentives, the hard stop policy group includes eleven states (AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, and OK), and the quality improvement initiative group has eleven states (AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT).

The DD identification strategy relies on the untestable assumption that treatment states would have similar trends to the control group if the policy had not been implemented. If trends are parallel in the pre-policy period, even if there is a difference in magnitude, I assume that differential changes in the post-period are driven by the policy, rather than inherent differences between regions. For each control group, I compare pre-policy trends in EEDs and low-risk c-sections to trends in the treatment states. All groups have statistically similar trends.

I estimate the impact of the Medicaid payment policy using the following equation:

$$Y_{mst} = \beta_0 + \beta_1 \cdot \text{Treat}_s + \beta_2 \cdot \text{Post}_t + \beta_3 \cdot \text{Treat}_s \cdot \text{Post}_t + T_t + \vartheta \cdot Z_m + \mu \cdot V_s + \varepsilon_{mst} \quad [6]$$

In [6], β_3 is the coefficient of interest, and it represents the aggregate effect of the Medicaid payment policy. Y_{mst} is the expected value of the outcome, discussed further below. It is indexed by State-MSA *m*; in state *s*; at time *t*, which is representative of pre/post policy implementation. Treat is a binary variable that denotes the presence of the Medicaid payment policy, and Post is a binary variable that indicates the policy post-period. V_s and T_t are state and year fixed-effects, respectively.

Z_m is a vector of time-varying State-MSA-level controls, which account for maternal characteristics, healthcare factors, and demographic and economic variables that may influence physician practice patterns. Maternal characteristics, drawn from the MarketScan data, include the percent of commercially insured mothers that are over 35 years old, the percent with a hospital length of stay over four days (the number of days typically covered by the insurer), and cost sharing quartile bins (National Conference of State Legislatures, 2020). Healthcare factors, extracted from AHA, AHRF, NPDB, and the MarketScan databases, include hospital characteristics (percent of hospitals that are non-profit, percent of hospitals that provide obstetric services, beds per 1,000, and percent of patients that are insured by Medicaid), primary care practitioners per 1,000, and financial attributes (average price differential between commercial c-sections and vaginal deliveries, and malpractice risk, defined as the average obstetric-related malpractice payout). Demographic characteristics from the ACS include percent of the population with less than a high school education, percent of the population with more than a college education, and percent of the population that is Black. Finally, economic characteristics from the ACS and AHRF include percent uninsurance, percent unemployment, and percent poverty.

I add dummy variables for the number of waves that a State-MSA appears in the data. I use wild bootstrapped standard errors, resampled 1,000 times and clustered at the State-MSA level. This accounts for covariance in geographic areas over time by multiplying the residual from each observation in a given cluster with a random variable that mimics the correlation in each cluster (Mackinnon, 2015; Roodman, Nielsen, Webb, & Mackinnon, 2018). This method is appropriate when there are few treated clusters and residuals are heteroskedastic, leaving traditional standard error approaches prone to over-rejection (Cameron & Miller, 2015; Colin Cameron, Gelbach, &

Miller, 2008; Conley & Taber, 2005) (Figure 2-3 and Figure 2-4). All models are estimated using Ordinary Least Squares.

The primary outcome is the all-payer EED rate, defined as the percent of patients in the State-MSA-quarter with elective vaginal deliveries or elective cesarean births between ≥ 37 and < 39 weeks' gestation completed, excluding individuals with conditions justifying elective delivery prior to 39 weeks. Justifiable conditions include comorbidities in the prenatal period (e.g. hypertension, diabetes, eclampsia, breech, and fetal abnormalities), and pregnancy complications (e.g. prolonged labor, fetal distress, or premature rupture of membranes) (Glantz, 2005). I measure this rate using the Joint Commission's Perinatal Care-01 (PC-01) methodology. Hospital Compare mandated that all hospitals with annual births totaling 1,100 or more submit PC-01 for public reporting, beginning January 1, 2013 (Joint Commission, 2019). I rely on this measure instead of commercial claims because International Classification of Diseases, Ninth Revision (ICD-9) codes lack the granularity needed to properly measure gestational age for EEDs. ICD-9 provides a single code for 37 or more completed weeks gestation, making it impossible to identify early-term births that occur between 37 and 39 weeks. The secondary outcome in this study is low-risk c-sections, defined as the percent of nulliparous women with a term, singleton baby in a vertex position delivered by c-section. I follow the methodology developed by the Agency for Healthcare Research and Quality (AHRQ), using Inpatient Quality Indicator (IQI) 33 (Agency for Healthcare Quality and Research, 2016). PC-01 and IQI 33 aggregate a rolling four-quarter measure rate, to ensure an adequate denominator size. Both measures are endorsed by the NQF as a consensus standard for hospital care.

Since PC-01 is all-payer, I cannot disentangle the direct result within Medicaid versus the spillover to privately insured patients. I conduct additional analyses to gain insight into this question by examining whether effects increase in geographic areas with a high share of Medicaid patients, split at the median level in the pre-policy period. I expect that if effects spilled over to the commercial sector, then the reduction in EEDs would remain constant across different levels of the policy exposure, measured by the Medicaid share. There is also some overlap in outcomes, as early elective c-sections are captured by both measures, but I cannot identify the degree of overlap. This can mask important effects if early term c-sections are decreasing, while full-term c-sections are increasing. To account for this, I avoid double-counting early inductions and c-sections using Allen and Grossman's (2019) approach, which treats inductions as the "absorbing state." In other words, if a patient's claim indicates both an induction and a c-section, I assume that the induction occurred first and exclude these observations from the low-risk c-section measure. There is little evidence to suggest that this is a substantial concern (Allen & Grossman, 2019).

I pursue a variety of robustness checks. First, I repeat the analyses using multiple group propensity score weights proposed by Stuart et al. (2014). Propensity scores aim to reduce selection bias by constructing a control group similar to the treatment, weighting units on a set of observed covariates. This reduces extrapolation of the counterfactual and aggregates a large number of confounders into a simple scalar (Stuart et al., 2014). However, when treatment and comparison states come from different underlying populations, propensity scores can introduce regression to the mean bias by selecting the "unusual" individuals from each of the two populations (e.g., those with lower values in the treatment group and higher values in the control group) (Daw & Hatfield, 2018); this introduces a tradeoff between improved covariate balance (with propensity scores) and regression to the mean bias (without propensity scores). I use an unweighted approach for the main analysis, but run supplemental models with multiple group propensity score weights.

Second, I employ an event study to determine if results are robust to standardized implementation timing. I also run sensitivity tests for alternate samples and outcomes. In the third sensitivity analysis, I re-run analyses using alternate treatment groups by dropping one treatment state at a time. Fourth, I re-estimate the models using individual quarterly outcome rates, rather than rolling four-quarter rates. Fifth, I re-run analyses excluding observations that did not report outcomes for all waves of data. Sixth, I test whether results are sensitive to inclusion of non-metropolitan areas. Finally, I re-run models with a smaller unit of analysis to determine whether inferences are stable with a larger sample size.

2.5. Results

2.5.1. Descriptive Statistics and Validity of Study Design

Table 2-1 summarizes State-MSA characteristics in treatment and control states, before and after implementation of the Medicaid payment policy. Differences in maternal and healthcare characteristics are relatively small in the pre-policy period. Prior to the Medicaid payment policy, treatment states had fewer mothers over 35 years old, more births with LOS over 4 days, and greater average cost sharing, compared to the main control group. Treatment states also had more Medicaid patients, fewer non-profit hospitals, fewer hospitals that provide obstetric care, lower density of primary care physicians, and a smaller price differential between c-sections and vaginal deliveries among the commercial population. Hospital beds per 1,000 were consistent between the groups. Demographic and economic differences were greater. On average, the population in treatment states was less educated, and more likely to be Black, uninsured, unemployed, or impoverished compared to the main control group. Differences in the other control groups are similar for most variables; however, I observe a few contrasts. The quality improvement control group was marginally less educated than the treatment group. The pay-for-performance control group had slightly fewer mothers over 35 years old and fewer births with LOS over four days. Quality improvement and pay-for-performance groups had a higher share of Medicaid patients; these groups, along with the hard stop policy group, also had fewer hospital beds per 1,000.

There is little evidence of differential changes in State-MSAs after the Medicaid payment policy is implemented. The gap in maternal age, as well as price differential between c-sections and vaginal deliveries, decreased, while differences in the Medicaid share increased. Average cost sharing rose across all intervention groups, while the uninsurance rate dropped. Other maternal, healthcare, demographic, and economic characteristic evolved similarly over time.

As discussed earlier, the DD approach relies on the assumption that trends in treatment and control states would evolve similarly without the Medicaid nonpayment policy. Although I cannot test this directly, I analyze pre-implementation trends to assess validity of the analysis. Figure 2-1 and Figure 2-2 plot unadjusted quarterly means of EEDs and low-risk c-sections, respectively. The red lines represent outcome trends in the treatment group and the blue lines represent outcome trends in each control group. Visual inspection suggests that EEDs follow similar pre-policy trends in treatment and control groups (Figure 2-1), while other outcomes are noisier (Figure 2-2). To verify that pre-trends are statistically similar, I run formal tests to assess differential pre-trends (Table 2-5). I use the main model specification, but limit inclusion to the pre-policy period. The coefficient of interest is the interaction between treatment and a linear quarter-year time trend. I find no statistically significant differences in trends, with the magnitude of all differences below 1%. These results are consistent with a valid identification strategy. But, since I cannot observe the counterfactual, it is important to apply a critical lens to the theoretical considerations.

One concern is that controlling for factors that are theoretically associated with higher levels of the outcome at baseline may mask important differences in trends (Kahn-Lang & Lang, 2018; Roth, 2019). For example, treatment states have fewer non-profit hospitals, which may contribute to lower initial levels of quality. This context is helpful for highlighting variation across groups, and provides a possible explanation for visual differences in pre-trends. Parallel counterfactual trends might be violated if the "common shocks" assumption is not upheld (e.g. if one group has a policy shock not experienced by the other group) (Dimick & Ryan, 2014; Ryan, Burgess, & Dimick, 2015). Similarly, part of the parallel counterfactual trends assumption is that covariates experience similar changes in magnitude and direction across treatment and control groups. To my knowledge, there are no obvious time-varying unobservable characteristics that would confound the policy effect. Medicaid expansion did not affect eligibility for pregnant women, who were already eligible up to 200% of the federal poverty line (Carroll et al., 2018). Synchronous federal policy (e.g. Choosing Wisely guideline advising against EEDs) may contribute to declining trends, but I expect these policies contribute equally to both groups (Figure 2-5). Based on these factors, I infer that counterfactual trends are parallel, further validating the study design.

2.5.2. Effect of Medicaid Payment Policy on Early Elective Deliveries

I first assess the effect of the Medicaid payment policy on EEDs, the primary target of the policy. Regression estimates of [6] are displayed in Table 2-2. Relative to the main control group, I find that EEDs decreased by 3.35% after the policy was implemented (95% Confidence Interval (CI): -6.69, -0.12). I observe larger effects in the treatment group compared to the hard stop policy group and the pay-for-performance group, with EEDs declining by 3.92% (95% CI: -6.98, -0.82) and 3.58% (95% CI: -6.73, -0.61), respectively. I estimate a smaller, insignificant decrease in EEDs compared to quality improvement programs. These results show that Medicaid nonpayment achieved its intended goal of reducing statewide all-payer EEDs, with a stronger response observed in comparison to a voluntary non-financial incentive and a financial bonus, but not an education-driven collaborative approach.

Next, I study whether the effect of the Medicaid payment policy varies by policy exposure, or the share of Medicaid patients (Table 2-3). I find no evidence of differential changes in EEDs between areas with a higher versus lower proportion of Medicaid patients. When comparing the Medicaid payment policy to the main control group, there is an insignificant decline in EEDs of 1.17% in areas with fewer Medicaid patients (95% CI: -5.69, 7.57). No comparison groups have significant differences. Since changes in all-payer EEDs do not vary by the share of Medicaid patients, my analysis suggests that both commercial and Medicaid patients were affected by the Medicaid payment change, and is suggestive of a spillover to the privately insured.

To understand what is driving the decrease in EEDs, I estimate whether effects of the Medicaid payment policy on EEDs vary across providers and areas with different financial and reputational characteristics. To do so, I modify [6] by adding an interaction term that multiplies treatment, post-policy period, and a binary variable for the characteristic of interest.

I first examine whether there are heterogeneous effects by the commercial price differential between c-sections and vaginal deliveries in a given geographic area (Table 2-3). A lower price difference means a relatively greater profit for physicians who substitute a low-value service (an EED) with a high-value service (a full-term vaginal delivery). This presents a larger incentive to reduce EEDs if physicians consider tradeoffs in treatment decisions across the entire choice set. Consistent with this hypothesis, I find a greater effect of the Medicaid payment policy in areas with a lower price differential, for all comparison groups. This difference is marginally significant between treatment states and pay-for-performance states, as areas with a lower price differential exhibit a 4.67% greater decline in EEDs compared to areas with a higher price differential (95% CI: -0.80, 10.15). Effects relative to other control groups are not significant.

I also assess whether effects of the Medicaid payment policy vary between areas with a higher versus lower proportion of for-profit hospitals (Table 2-3). I find the reduction in EEDs to be greater in areas with more for-profit hospitals among treatment states. This difference is significant in treatment states compared to states with hard stop policies (β: 6.53%; 95% CI: -13.13, -0.02) and pay-for-performance initiatives (β: 7.65%; 95% CI: -13.74, -1.39). These results are consistent with my conceptual model, which suggests that physicians in for-profit hospitals may be more responsive to changes in service profitability compared to physicians in non-profit hospitals.

2.5.3. Effect of Medicaid Payment Policy on Low-Risk C-Sections

In Table 2-4, I study the effect of the Medicaid payment policy on supply of low-risk c-sections, to explore whether physician-induced demand is present in the commercial sector. If physicians are inducing demand to recoup profits lost from EEDs, I would expect to see an increase in low-risk c-sections in treatment states. I find little evidence of changes in low-risk c-sections. There is a small, statistically insignificant increase of 1.21% in low-risk c-sections in the treatment group compared to the main control group (95% CI: -1.36, 3.66). Since low-risk c-sections are declining across all groups, I can interpret this as a smaller decrease within treatment states.

Changes in low-risk c-sections relative to the other comparison groups are insignificant and small in magnitude (at or below 0.5%).

I further test whether there is evidence of induced demand by exploring whether there is a larger increase in low-risk c-sections in areas with higher policy exposure (measured by the share of Medicaid patients), and in areas where c-sections are more profitable (measured by the price differential between c-sections and vaginal deliveries in the commercial sector). If physicians replace the volume of EEDs with low-risk c-sections, this suggests a positive income effect. Similar to the EED analysis, I add a three-way interaction multiplying the treatment, post-period, and a binary indicator for high policy exposure or high profitability of c-sections [6]. I find no evidence of induced demand. Results in Table 2-4 indicate that low-risk c-sections decline to a greater extent in the treatment group in areas with a higher share of Medicaid patients, but these differences are small and insignificant (below 1.0%). There is also a small, but insignificant, increase of low-risk c-sections in areas with a high price difference between c-sections and vaginal deliveries. This suggests that the policy's effect on low-risk c-sections does not vary by policy exposure or c-section profits, which would be expected under the demand inducement hypothesis.

2.5.4. Robustness Checks

In this section, I assess robustness of the results. First, I repeat the analyses using multiple group propensity score weights proposed by Stuart et al. (2014). In Table 2-7, I measure whether covariate balance improved with propensity score weights using the standardized mean difference (SMD) between treatment and control means. I construct multiple group propensity score weights separately in the pre- and post-periods, using logistic regression models. I observe improved balance with propensity scores. SMDs are below the recommended threshold of 0.25 for the majority of covariates (Stuart, Lee, & Leacy, 2013). Results are robust to propensity score weighting, with treatment states experiencing a 3.27% reduction in EEDs relative to the main control (95% CI: -7.16, 0.89), versus 3.35% in the main analysis (Table 2-6).

Second, I repeat analyses as an event study to determine if results are robust to standardizing timing of policy implementation across states. I follow Callaway and Sant'Anna (2020), which measures the "group-time average treatment effect" by adding interaction terms between treatment, post-period, and year of treatment implementation (Callaway & Sant'Anna, 2020). For control groups, I add a binary indicator for "never-treated." Results were generally consistent with the main analysis, except with significantly larger spillovers in the event study, ranging from 14.35% to 20.74% (Table 2-12; Figure 2-6 and Figure 2-7). The discrepancy in the event study result is likely driven by high pre-treatment EED rates in MS, the only state to enact the Medicaid payment policy in 2015 (rather than 2014). Since this event study approach assigns equal weight to each group-time effect, MS receives disproportionate weight, thereby inflating point estimates (Sun & Abraham, 2020). Thus, I rely on the main analysis to make inferences on spillover magnitude.

Next, I re-run analyses using alternate treatment groups, each with a treatment state omitted to assess whether a single state disproportionately influenced results. I find that estimated reductions in EEDs range from 2.34% to 4.46% in the main control group, although models without GA or MS were no longer significant. All other comparisons had estimates comparable in direction and magnitude to the main analysis (Table 2-9). Fourth, I re-estimate the models using individual quarterly rates for EEDs and low-risk c-sections, instead of a rolling four-quarter measure period. I find that models are not sensitive to the alternate measurement specification (Table 2-8). Fifth, I re-run the models with non-metropolitan State-MSAs, finding that results are robust to inclusion of these observations (Table 2-10). Sixth, I re-run analyses excluding all State-MSAs that did not report outcomes for all waves of data (Table 2-11). I observe a 2.83% decline in EEDs compared to the main control group (95% CI: -6.51, 0.88). This suggests that missing values of the outcome, driven by variation in public reporting requirements over time, did not systematically impact results. Finally, results were not sensitive to using hospital-quarter or patient, rather than State-MSA, as the unit of analysis. Hospital and patient level analyses increased the sample size to 3,102 and 91,561, respectively, and led to comparable inferences (Table 2-13). Robustness checks showed no significant changes in low-risk c-sections across all specifications, and the magnitude and direction remained fairly constant.

2.6. Discussion

In this paper, I study the indirect effects of a Medicaid nonpayment incentive implemented between January 1, 2014 and January 1, 2015 for EEDs, compared to other financial and non-financial incentives. I compare how the Medicaid payment policy impacted use of low-value childbirth deliveries in the commercial market relative to states with other Medicaid policies, including: no policy aimed at reducing EEDs, a hard stop policy, a quality improvement collaborative, or a pay-for-performance reform.

First, I explore whether the Medicaid payment policy prompted a reduction of EEDs among privately insured patients. I find that the Medicaid payment policy in GA, IN, MO, and MS led to a 3.35% decline in all-payer EEDs compared to states with no EED policy. The Medicaid payment policy also reduced all-payer EEDs by 3.92% and 3.58% compared to states with a hard stop policy and pay-for-performance program, respectively. Effects did not vary between geographic areas with a higher versus lower share of Medicaid patients. This provides suggestive evidence that both Medicaid and commercial patients were impacted by the policy, and indicates a positive spillover to the privately insured population.

Next, I assess whether there was evidence of physician-induced demand by examining if there was an increase in low-risk c-sections. I do not find statistically significant changes in the rate of privately insured low-risk c-sections. Effects do not increase in areas with high policy exposure or in areas where c-sections are more profitable, suggesting that physicians do not substitute the volume of EEDs with low-risk c-sections. Taken together, this offers no evidence of physician-induced demand, which would be expected in areas where the decline in profits is high and the time cost is low. Sensitivity tests support the main conclusions.

Finally, I explore whether the variation in spillovers was consistent with financial or reputational drivers. I find changes to be consistent with financial drivers. I test whether the decline in EEDs was greater in areas with a higher share of for-profit hospitals, which would be expected if physician respond more strongly to incentives that align with hospital objectives. I find a greater reduction in EEDs in areas with more for-profit hospitals in treatment states compared to hard stop and pay-for-performance groups. The magnitude of this effect is 6.53% and 7.65%, respectively. One explanation is that physicians in for-profit hospitals, motivated primarily by financial objectives, respond more strongly to changes in financial incentives than those in non-profits. In prior studies, for-profit hospitals demonstrated a more significant reaction to changes in the financial environment by offering a greater volume of profitable services (Dranove et al., 2017; Horwitz, 2005). As Medicaid EEDs become less profitable, it is possible that for-profit hospitals respond by reducing the overall provision of EEDs. Another explanation is that non-profit hospitals prioritize reputation, regardless of whether financial incentives are present, prompting a lesser response to the Medicaid payment change (Newhouse, 1970). This may point to barriers in changing physician practice patterns when objectives are not aligned with the hospital's. I also observe a larger decrease in EEDs in areas with a lower price differential between c-sections and vaginal deliveries in the commercial sector. This is notable because a lower price difference between these services means that high-value substitutes are relatively profitable. This suggests that physicians consider tradeoffs across the entire choice set when making treatment decisions.

As several states and payers continue to debate payment policies to reduce low-value care, the analysis suggests that nonpayment incentives can be successful on a large scale. The magnitude of the results, however, indicates that Medicaid reimbursement may have a relatively modest spillover to commercially insured patients, compared to the direct effect within Medicaid. The estimate of a 3.35% decline in EEDs is significantly smaller than those observed in Medicaid populations in Texas (14%) and South Carolina (16.6%) (Allen & Grossman, 2019; Dahlen et al., 2017). The effect size is comparable to evaluations of the non-Medicaid population in Texas, which also found significantly larger effects on Medicaid versus non-Medicaid EEDs. The difference in Texas was attributed to the Medicaid population facing both hard stop and reimbursement changes, while the non-Medicaid population faced only the hard stop policy, rather than a direct spillover (Byanova, 2015). In contrast, I do not find effects of the Medicaid payment policy on commercial c-sections. Prior work found a 13.1% increase in non-Medicaid c-sections in Texas, with effects concentrated in hospitals with more Medicaid births (Byanova, 2015). This contrast may be due to differences in how I measure c-sections; the prior study examines the total c-section rate, while I focus on low-risk c-sections. Low-risk c-sections provide a more precise measure of demand inducement by capturing overuse of a low-value service. Since c-sections can be appropriate for risky births, focusing on the total c-section rate limits the ability to make inferences about whether treatment intensity is higher than optimal (Goer, Romano, & Sakala, 2012). The results contribute to a growing empirical literature on financial incentives and physician behavior, showing that incentives can lead to positive spillovers across payers.

These results have several implications for efforts to expand non-FFS payment reforms. First, this analysis indicates that discontinuing payment for a low-value service has potential to reduce unwanted physician behaviors, compared to the status quo. This is notable because Medicaid nonpayment is a relatively large penalty, as most value-based programs offer little to no downside risk. This work also suggests that financial nonpayment can yield greater improvements in statewide quality than a range of other financial and non-financial incentives. The largest gains among treatment states were made relative to hard stop policies. Hard stop initiatives are voluntary and strategies vary between hospitals. In contrast, the Medicaid payment policy rollout is standardized within states, and participation is mandatory. Thus, mandatory programs may have greater potential to engage providers with the highest levels of low-value care. Literature focused on mandatory incentive programs is limited; in general, studies conclude that spending reductions are greater in voluntary programs than in mandatory ones, but effects on quality are inconclusive (Dummit et al., 2016; Finkelstein, Ji, Mahoney, & Skinner, 2018; Liao, Sommers, & Navathe, 2018; Navathe et al., 2018, 2017). This study offers preliminary evidence that low-value care may be reduced more effectively under mandatory, rather than voluntary, policies. Further, these results offer initial evidence that imposing clear, coordinated goals across health systems may lead to larger declines in low-value treatment.

Second, given significant reductions in EEDs due to Medicaid nonpayment compared to pay-for-performance payment, financial penalties may lead to greater spillovers than bonuses. These results are consistent with prospect theory in behavioral economics, which posits that providers value gains and losses of the same magnitude asymmetrically, and will respond more strongly to penalties to avoid financial losses (Kahneman & Tversky, 1979). Empirical evidence supports this notion, but this is the first study to explore spillovers across payers under this framework (Rizzo et al., 2002). Finally, Medicaid nonpayment for EEDs is equally as successful as quality improvement programs in reducing EEDs. Since quality improvement targets behavior changes across all payers, this suggests that Medicaid-specific payment reform may lead to stronger spillovers to other payers. It may also highlight how a non-financial incentive, with the multi-pronged approach of education, stakeholder engagement, and performance measurement, can achieve similar results to Medicaid payment reform.

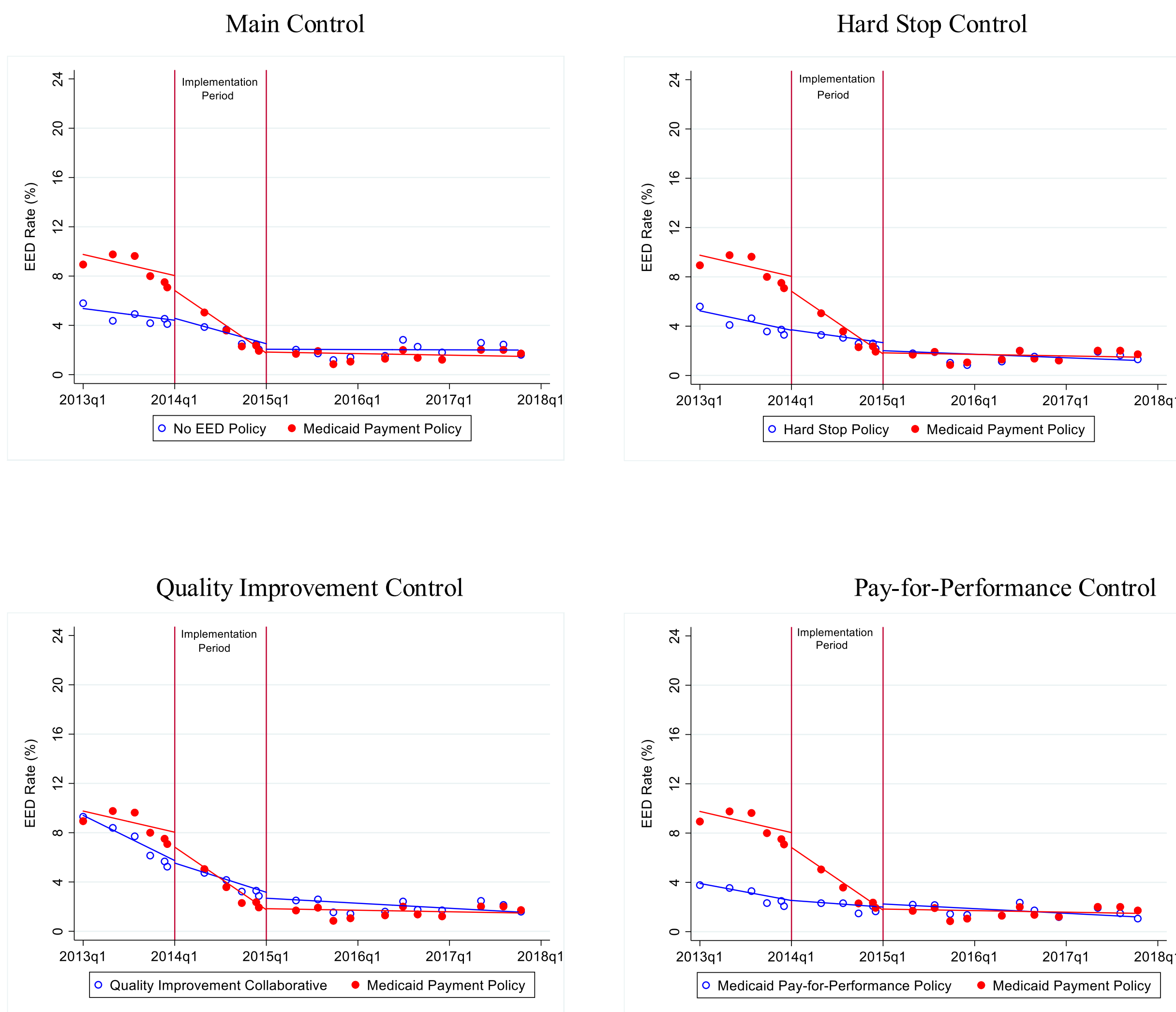
This analysis has several limitations and suggests potential avenues for future work. First, I cannot identify physicians in the Hospital Compare or MarketScan databases. This limits the ability to attribute effects to changes in physician, rather than hospital, behavior. Further, since regional distribution of for-profit hospitals is not uniform, it is possible that other geographic factors (e.g. hospital concentration, physician-hospital integration), may contribute to variation in spillovers (American Hospital Association, 2018; Fulton, 2017; Health Care Cost Institute, 2017; The Henry J. Kaiser Family Foundation, 2018). It will be useful for future work to address this gap and examine these effects at the physician level. Exploring this area would enable a better understanding of physician characteristics that drive variation in policy effects, and the potential mechanism behind behavioral changes.

Second, I am only able to identify all-payer, rather than commercial insurance-specific, EED rates. This limits the ability to make inferences about spillovers to the privately insured. I argue that since policy effects do not vary by the share of Medicaid patients, this suggests a possible effect on non-Medicaid births. To be certain, additional research focusing on commercial-specific EED rates, is needed. Third, I focus on spillovers of Medicaid payment reform in childbirth, so results may not be generalizable to other clinical areas. I argue that perinatal care has characteristics that reduce potential for effects to stem from confounders, which can increase application of results to other settings. Additional work is needed to strengthen this claim, especially given the small sample size. Next, the DD design rests on the assumption that no unobserved factors contribute to the observed effect. I address this concern with several robustness checks. However, minimal availability of data in the pre-policy period weakens interpretation of results. Finally, treatment states have lower initial levels of quality, leading to questions of whether federal guidelines, rather than the Medicaid policy, drive the observed effect. I simulate the expected effect of the guidelines (in the absence of the Medicaid policy) by assuming that treatment states follow an analogous trend to control states with no EED policy in the post-period (Figure 2-5). The figures show a significantly smaller expected decline in EEDs relative to the observed effect with the Medicaid payment policy, suggesting that spillovers stem from Medicaid nonpayment. The broader empirical literature has shown no strong evidence of significant improvements from Choosing Wisely guidelines (Grimshaw et al., 2020; Hilal & Munoz, 2020), but more research is needed to evaluate how these guidelines work in tandem with financial incentives targeting low-value care.

Non-FFS payment reforms are becoming increasingly salient, but there are remaining questions about how to design incentives that promote high-value care and generate cost savings. I present evidence that Medicaid incentives can be effective in reducing low-value care in the commercial sector, without prompting unintended consequences such as physician-induced demand. While the results are specific to perinatal care, this study provides general guidance about financial and non-financial design. Continuing to build an understanding of these incentives and their indirect effects, especially surrounding variation in success, is imperative for future work.

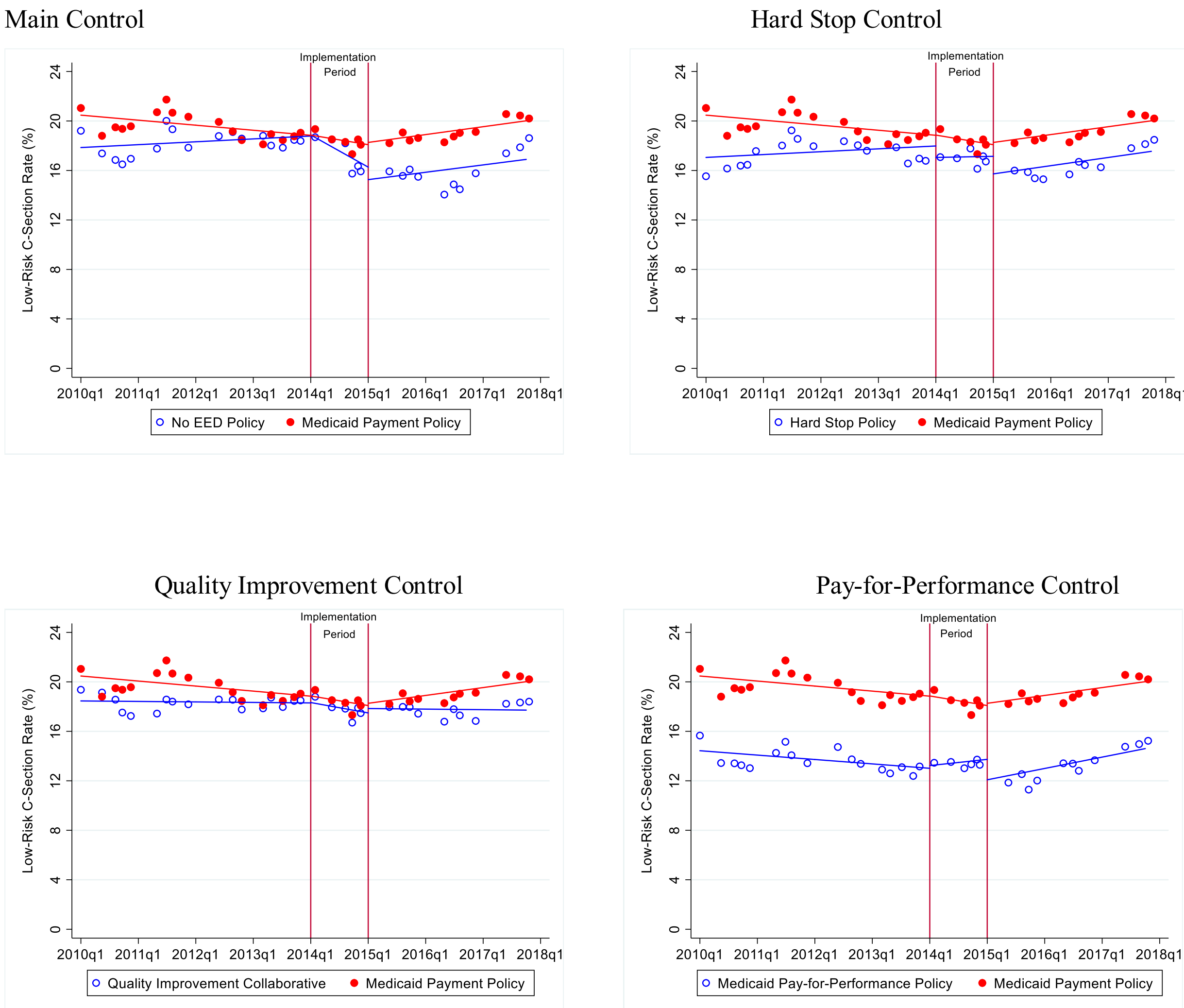
2.7. Figures

Figure 2-1 Unadjusted Trends in Rates of Early Elective Deliveries in Treatment and Control States



Notes: Sample estimates for EEDs are from Hospital Compare in 2013-2017. Data points are unadjusted, quarterly means. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI).

Figure 2-2 Unadjusted Trends in Rates of Low-Risk Cesarean Sections in Treatment and Control States



Notes: Sample estimates for Low-risk c-sections from Truven MarketScan claims, using data from births between 2010-2017. Data points are unadjusted, quarterly means. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI).

2.8. Tables

Table 2-1 Summary Statistics of Treatment and Control States, Pre and Post Medicaid Payment Policy Implementation

	Pre-Medicaid Policy (2013)					**Post-Medicaid Policy (2015-2017)**				
	Treatment Sample	Control Sample				Treatment Sample	Control Sample			
	Medicaid Policy	Main	Hard Stop	QI	P4P	Medicaid Policy	Main	Hard Stop	QI	P4P
Maternal Characteristics										
% Maternal Age 35+	10.86	13.35	13.05	14.18	10.73	11.80	13.54	12.78	13.75	12.91
% LOS > 4	3.64	3.27	4.26	3.99	3.42	4.37	3.57	4.52	4.96	3.35
Avg. % Cost Sharing	15.57	15.23	14.57	11.76	14.05	16.01	17.07	17.33	13.50	14.48
Healthcare Characteristics										
% Medicaid	15.38	14.79	14.30	17.26	15.41	15.80	14.21	16.68	19.96	17.05
% Hospitals Non-Profit	48.49	56.59	60.83	53.05	66.23	49.54	54.55	58.51	51.66	68.37
% Hospitals Provide OB Services	39.98	43.85	52.82	48.54	62.16	39.66	39.39	50.66	72.65	58.42
Beds per 1,000	2.90	2.91	2.15	2.11	1.72	2.96	3.05	2.35	2.16	1.73
PCPs per 1,000	0.68	0.85	0.82	0.72	0.81	0.67	0.86	0.82	0.73	0.79
Avg. Price Differential										
< $0	1.48	0.93	3.24	3.88	1.71	3.83	3.34	2.68	4.52	2.23
$0 - $5,000	70.37	42.06	55.76	51.04	31.62	60.53	49.85	55.43	47.98	35.47
≥ $5,000	28.15	57.01	41.01	45.07	66.67	35.65	46.81	41.89	47.49	62.29

Demographic Characteristics										
% < HS Education	13.74	10.02	10.58	13.82	9.80	12.64	8.96	9.81	13.00	8.88
% > College Education	25.00	30.81	29.67	26.06	28.99	27.08	32.82	30.94	27.72	30.95
% Population Black	22.04	8.56	12.70	11.86	3.51	22.13	8.59	12.67	11.98	3.62
Economic Characteristics										
% Uninsured	18.52	14.81	13.97	16.94	13.88	13.20	10.28	8.66	9.60	7.17
% Unemployed	7.77	6.08	7.08	8.61	7.17	5.01	4.21	4.55	6.09	4.58
% Poverty	18.23	13.61	16.07	16.76	13.26	16.09	12.25	14.24	15.00	11.66
State-MSA-Quarters	135	107	278	335	117	418	329	857	1,017	358
# of States	4	8	11	11	3	4	8	11	11	3

Notes: Sample estimates are from the Hospital Compare Database in 2013-2017. Maternal characteristics are from the Truven MarketScan commercial claims database, using data from births during the study period. Healthcare characteristics are from the AHA Annual Survey, AHRF, and the NPDB. Average price differential represents the mean difference in reimbursement between c-sections and vaginal deliveries among births in the Truven MarketScan data. Demographic and economic characteristics are from the U.S. Census ACS. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI).

Table 2-2 Spillover Effects of Medicaid Payment Policy on Low-Value Care Outcomes

	Main	Hard Stop	QI	P4P
Early Elective Deliveries				
Treatment * Post	-3.35**	-3.92**	-1.44	-3.58**
	(-6.69, -0.12)	(-6.98, -0.82)	(-4.60, 1.87)	(-6.73, -0.61)
N	989	1,688	1,905	1,028
Dependent Variable Mean: % (SD)				
Pre: Treatment Mean	9.01 (9.50)	9.01 (9.29)	9.01 (9.29)	9.01 (9.29)
Pre: Control Mean	4.96 (4.83)	4.60 (7.91)	8.01 (8.45)	3.39 (3.20)
Post: Treatment Mean	1.67 (1.90)	1.67 (1.90)	1.67 (1.90)	1.67 (1.90)
Post: Control Mean	2.03 (2.77)	1.61 (1.94)	2.12 (2.12)	1.73 (1.59)
Low-Risk C-Sections				
Treatment * Post	1.21	0.11	0.01	0.50
	(-1.36, 3.66)	(-1.60, 1.86)	(-1.71, 1.74)	(-1.86, 2.94)
N	1,773	3,021	3,421	1,827
Dependent Variable Mean (SD)				
Pre: Treatment Mean	19.80 (6.67)	19.80 (6.67)	19.80 (6.67)	19.80 (6.67)
Pre: Control Mean	18.17 (9.23)	17.59 (7.81)	18.38 (6.20)	13.76 (7.05)
Post: Treatment Mean	19.08 (6.36)	19.08 (6.36)	19.08 (6.36)	19.08 (6.36)
Post: Control Mean	16.19 (6.84)	16.61 (6.61)	17.92 (6.91)	13.36 (5.27)

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013-2017. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 2-3 Variation in Spillover Effects of Early Elective Deliveries by Financial Market Characteristics

	Main		Hard Stop		QI		P4P	
High vs. Low Share of Medicaid Patients								
Treat * Post * High Medicaid	1.17 (-5.69, 7.57)		-1.74 (-8.12, 4.79)		-4.74 (-11.49, 1.74)		-0.81 (-7.26, 5.60)	
N	989		1,688		1,905		1,028	
Dependent Variable Mean: %								
High Medicaid	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
Treatment	10.29	1.94	10.29	1.94	10.29	1.94	10.29	1.94
Control	5.81	1.76	4.02	1.72	6.71	1.98	3.31	1.58
Low Medicaid								
Treatment	8.18	1.39	8.18	1.39	8.18	1.39	8.18	1.39
Control	4.46	2.18	5.21	1.51	10.29	2.37	3.45	1.84
High vs. Low Price Difference: Commercial C-Sections vs. Vaginal Deliveries								
Treat * Post * High Price Diff.	0.42 (-7.95, 7.99)		2.97 (-2.96, 9.03)		1.20 (-4.80, 7.35)		4.67* (-0.80, 10.15)	
N	989		1,688		1,905		1,028	
Dependent Variable Mean: %								
High Price Difference	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
Treatment	6.71	1.40	6.71	1.40	6.71	1.40	6.71	1.40
Control	4.04	2.05	3.65	1.63	6.48	2.05	3.30	1.60
Low Price Difference								
Treatment	11.07	1.86	11.07	1.86	11.07	1.86	11.07	1.86
Control	8.15	1.96	5.56	1.59	9.92	2.21	3.99	2.55
High vs. Low % For-Profit Hospitals								
Treat * Post * High % For-Profits	-3.88 (-10.43, 2.94)		-6.53** (-13.13, -0.02)		-2.50 (-9.10, 3.98)		-7.65** (-13.74, -1.39)	
N	989		1,688		1,905		1,028	
Dependent Variable Mean: %								
High % For-Profit Hospitals	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
Treatment	13.32	1.74	13.32	1.74	13.32	1.74	13.32	1.74
Control	6.70	1.85	5.05	1.65	10.52	2.25	1.97	1.14

Low % For-Profit Hospitals								
Treatment	5.49	1.59	5.49	1.59	5.49	1.59	5.49	1.59
Control	3.84	2.15	4.39	1.60	6.00	2.02	3.66	1.85

Notes: Sample estimates for EEDs are from Hospital Compare in 2013-2017. Table cells include DDD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Stratification variables are determined by a threshold of ≥ median in the pre-policy period (2010-2013). Price differential represents the mean difference in reimbursement between c-sections and vaginal deliveries among births in the Truven MarketScan data. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 2-4 Variation in Spillover Effects of Low-Risk Cesarean Sections: An Exploration of Physician-Induced Demand

	Main		Hard Stop		QI		P4P	
High vs. Low Share of Medicaid Patients								
Treat * Post * High Medicaid	-0.65 (-5.14, 3.82)		-0.59 (-4.10, 2.70)		-0.51 (-3.73, 2.85)		-0.33 (-5.15, 4.52)	
N	1,773		3,021		3,421		1,827	
Dependent Variable Mean: %								
High Medicaid	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
Treatment	21.66	20.92	21.66	20.92	21.66	20.92	21.66	20.92
Control	19.34	17.25	17.56	16.82	18.11	17.52	14.81	14.77
Low Medicaid								
Treatment	17.81	17.24	17.81	17.24	17.81	17.24	17.81	17.24
Control	17.46	15.59	17.63	16.40	18.89	18.64	12.95	12.29
High vs. Low Price Difference: Commercial C-Sections vs. Vaginal Deliveries								
Treat * Post * High Price Diff.	0.08 (-5.94, 5.68)		1.12 (-2.33, 4.44)		-0.68 (-3.82, 2.57)		1.16 (-9.74, 6.93)	
N	1,773		3,021		3,421		1,827	
Dependent Variable Mean: %								
High Price Difference	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*	*Pre*	*Post*
Treatment	18.58	17.91	18.58	17.91	18.58	17.91	18.58	17.91
Control	17.94	16.13	16.15	14.53	17.85	17.71	14.09	13.63
Low Price Difference								
Treatment	20.70	19.94	20.70	19.94	20.70	19.94	20.70	19.94
Control	18.92	16.42	19.08	18.72	19.00	18.19	11.56	11.62

Notes: Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DDD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Stratification variables are determined by a threshold of ≥ median in the pre-policy period (2010-2013) to reduce endogeneity. Price differential represents the mean difference in reimbursement between c-sections and vaginal deliveries among births in the Truven MarketScan data. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

2.9. References

Agency for Healthcare Quality and Research. (2016). *AHRQ - Quality Indicators Technical Specifications*. Retrieved from https://www.qualityindicators.ahrq.gov/Modules/IQI_TechSpec_ICD10_v60.aspx

Agrawal, A. (2017). *Case Study: Wisconsin's Obstetric Medical Home Program Promotes Improved Birth Outcomes*. Retrieved from www.nashp.org

Alexander, D. (2015). *Does Physician Pay Affect Procedure Choice and Patient Health? Evidence from Medicaid C-section Use* *. Retrieved from http://scholar.princeton.edu/dalexand

Allen, L., & Grossman, D. (2019). The impact of voluntary and nonpayment policies in reducing early-term elective deliveries among privately insured and Medicaid enrollees. *Health Services Research*, 1–8. https://doi.org/10.1111/1475-6773.13214

American Hospital Association. (2018). AHA Data Products: Annual Survey. Retrieved August 23, 2019, from https://www.aha.org/data-insights/aha-data-products

Association of State and Territorial Health Officials. (2014). *Medicaid Quality Improvement Measures-Reducing Elective Deliveries in Washington State*. Retrieved from www.astho.org

Baicker, K., Chernew, M., & Robbins, J. (2013). *The Spillover Effects of Medicare Managed Care: Medicare Advantage and Hospital Utilization*. Retrieved from http://www.nber.org/papers/w19070.ack

Bain. (2017). *Front Line of Healthcare Report*. Retrieved from https://app.shorthand.com/export/5c882d8b4d4c4fdfa68b0b8621070094/

Barnett, M. L., Olenski, A., & Sacarny, A. (2020). Common practice: spillovers from Medicare on private health care. *NBER Working Paper Series*, (27270). Retrieved from http://www.nber.org/papers/w27270

Brownlee, S., Chalkidou, K., Doust, J., Elshaug, A. G., Glasziou, P., Heath, I., … Korenstein, D. (2017). Evidence for overuse of medical services around the world. *Www.Thelancet.Com*, *390*. https://doi.org/10.1016/S0140-6736(16)32585-5

Brownlee, S., Saini, V., & Cassel, C. (2014). When Less is More: Issues On Overuse in Healthcare. *Health Affairs*, *370*(15), 1454–1457. https://doi.org/10.1056/NEJMe1402388

Byanova, D. (2015). *Is It Worth the Wait? Early Elective Deliveries, Procedure Use, and Neonatal Health*. Retrieved from https://www.brown.edu/academics/economics/candidates/sites/brown.edu.academics.economics.candidates/files/JMP - Desislava Byanova.pdf

Callaway, B., & Sant'Anna, P. H. C. (2020). Difference-in-Differences with Multiple Time Periods. *Working Paper*, 1–45.

Cameron, A. C., & Miller, D. L. (2015). *A Practitioner's Guide to Cluster-Robust Inference*. Retrieved from http://cameron.econ.ucdavis.edu/research/Cameron_Miller_JHR_2015_February.pdf

Carroll, A. E. (2017). Why Is US Maternal Mortality Rising? *JAMA*, *318*(4), 321. https://doi.org/10.1001/jama.2017.8390

Carroll, C., Chernew, M., Fendrick, A. M., Thompson, J., & Rose, S. (2018). Effects of episode-based payment on health care spending and utilization: Evidence from perinatal care in Arkansas. *Journal of Health Economics*, *61*, 47–62. https://doi.org/10.1016/j.jhealeco.2018.06.010

Caughey, A. B., Sundaram, V., Kaimal, A. J., Cheng, Y. W., Gienger, A., Sarah Little, B. E., … Bravata, D. M. (2009). *Evidence Report/Technology Assessment Number 176 Maternal and Neonatal Outcomes of Elective Induction of Labor*. Retrieved from www.ahrq.gov

Centers for Medicare and Medicaid Services. (2016). Hospital Compare Database. Retrieved September 11, 2019, from https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html

Chandra, A., & Skinner, J. (2012). Technology Growth and Expenditure Growth in Health Care. *Journal of Economic Literature*, *50*(3), 645–680. https://doi.org/10.1257/jel.50.3.645

Chernew, M., Baicker, K., & Martin, C. (2010). *Spillovers in Health Care Markets: Implications for Current Law Projections*.

Choosing Wisely. (2013). *The American Congress of Obstetricians and Gynecologists' Recommendations on Elective Labor Inductions and Cesarean Sections*. Retrieved from http://www.choosingwisely.org/clinician-lists/american-college-obstetricians-gynecologists-scheduling-elective-labor-inductions-or-cesarean-deliveries/

Chou, Y.-J., Huang, N., Lin, I.-F., Deng, C.-Y., Tsai, Y.-W., Chen, L.-S., & Lee, C.-H. (2006). Do Physicians and Their Relatives Have a Decreased Rate of Cesarean Section? A 4-Year Population-Based Study in Taiwan. *Birth*, *33*(3), 195–202. https://doi.org/10.1111/j.1523-536X.2006.00104.x

Clemens, J., & Gottlieb, J. D. (2014). Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? *American Economic Review*, *104*(4), 1320–1349. https://doi.org/10.1257/aer.104.4.1320

Coey, D. (2015). Physicians' financial incentives and treatment choices in heart attack management. *Quantitative Economics*, *6*, 703–748. https://doi.org/10.3982/QE365

Colin Cameron, A., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427. https://doi.org/10.1162/rest.90.3.414

Colorado Medicaid. (2016). *2016 Hospital Quality Incentive Payment (HQIP) Program*.

Conley, C., & Taber, T. (2005). *Inference with "difference in differences" with a small number of policy changes*. Retrieved from http://www.nber.org/papers/T0312

Consumer Reports. (2017). *Most U.S. Hospitals' C-Section Rates Exceeding National Targets*. Retrieved from https://www.consumerreports.org/media-room/press-releases/2017/05/consumer_reports_analysis_most_us_hospitals_csection_rates_exceeding_national_targets/

Dahlen, H. M., Mccullough, J. Mac, Fertig, A. R., Dowd, B. E., & Riley, W. J. (2017). Texas Medicaid Payment Reform: Fewer Early Elective Deliveries And Increased Gestational Age And Birthweight. *Health Affairs*, *36*(3), 460–467. https://doi.org/10.1377/hlthaff.2016.0910

Daw, J. R., & Hatfield, L. A. (2018). Matching in Difference-in-Differences: between a Rock and a Hard Place. *Health Services Research*, *53*(6), 4111–4117. https://doi.org/10.1111/1475-6773.13017

Dimick, J. B., & Ryan, A. M. (2014). Methods for evaluating changes in health care policy: The difference-in-differences approach. *JAMA - Journal of the American Medical Association*, *312*(22), 2401–2402. https://doi.org/10.1001/jama.2014.16153

Dranove, D., Garthwaite, C., & Ody, C. (2017). How do nonprofits respond to negative wealth shocks? The impact of the 2008 stock market collapse on hospitals. *The RAND Journal of Economics*, *48*(2), 485–525. https://doi.org/10.1111/1756-2171.12184

Dummit, L. A., Kahvecioglu, D., Marrufo, G., Rajkumar, R., Marshall, J., Tan, E., … Conway, P. H. (2016). Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes. *JAMA*, *316*(12), 1267. https://doi.org/10.1001/jama.2016.12717

Ellis, R. P., & McGuire, T. G. (1986). Provider behavior under prospective reimbursement. Cost sharing and supply. *Journal of Health Economics*, *5*(2), 129–151. https://doi.org/10.1016/0167-6296(86)90002-0

Evans, R. G. (1974). Supplier-Induced Demand: Some Empirical Evidence and Implications. In *The Economics of Health and Medical Care* (pp. 162–173). https://doi.org/10.1007/978-1-349-63660-0_10

Finkelstein, A., Ji, Y., Mahoney, N., & Skinner, J. (2018). Mandatory Medicare Bundled Payment Program for Lower Extremity Joint Replacement and Discharge to Institutional Postacute Care. *JAMA*, *320*(9), 892. https://doi.org/10.1001/jama.2018.12346

Foo, P. K., Lee, R. S., & Fong, K. (2017). Physician Prices, Hospital Prices, and Treatment Choice in Labor and Delivery. *American Journal of Health Economics*, *3*(3), 422–453. https://doi.org/10.1162/ajhe_a_00083

Frank, R. G., Zeckhauser, R. J., Osher, S., Rosenthal, D., Schoenberg, M., & Kennedy, J. F. (2007). *Custom Made Versus Ready to Wear Treatments; Behavioral Propensities in Physician's Choices*. Retrieved from http://www.nber.org/papers/w13445

Fulton, B. D. (2017). Health Care Market Concentration Trends in the United States: Evidence and Policy Responses. In *Journal Health Affairs*. Health Affairs.

Glantz, J. C. (2005). Elective induction vs. spontaneous labor: Associations and outcomes. *Journal of Reproductive Medicine for the Obstetrician and Gynecologist*, *50*(4), 235–240.

Glied, S., & Zivin, J. G. (2002). How do doctors behave when some (but not all) of their patients are in managed care? *Journal of Health Economics*, *21*(2), 337–353. https://doi.org/10.1016/S0167-6296(01)00131-X

Goer, H., Romano, A., & Sakala, C. (2012). *Vaginal or Cesarean Birth: What Is at Stake for Women and Babies?* New York.

Goodman-Bacon, A. (2018). Difference-in-Differences with Variation in Treatment Timing. *National Bureau of Economic Research*. https://doi.org/10.3386/w25018

Grimshaw, J. M., grimshaw, J., patey, andrea, Kirkham, K. R., hall, amanda, dowling, shawn K., & Rodondi, nicolas. (2020). De-implementing wisely: developing the evidence base to reduce low-value care. *BMJ Qual Saf*, *29*, 409–417. https://doi.org/10.1136/bmjqs-2019-010060

Gruber, Jon, Kim, J., & Mayzlin, D. (1999). Physician fees and procedure intensity: the case of cesarean delivery. In *Journal of Health Economics* (Vol. 18). Retrieved from www.hcfa.gov

Gruber, Jonathan ;, & Owings, M. (1996). Physician financial incentives and cesarean section delivery. In *The Rand Journal of Economics; Spring* (Vol. 27). Retrieved from https://economics.mit.edu/files/12487

Health Care Cost Institute. (2017). *Healthy Market Index Interactive*. Retrieved from https://healthcostinstitute.org/research/hmi-interactive

Hilal, T., & Munoz, J. (2020). Choosing Wisely® in Hematology: Have We Made a Difference? *Current Hematologic Malignancy Reports*, *15*, 241–247. https://doi.org/10.1007/s11899-020-00593-2

Horwitz, J. R. (2005). Making profits and providing care: Comparing nonprofit, for-profit, and government hospitals. *Health Affairs*, *24*(3), 790–801. https://doi.org/10.1377/hlthaff.24.3.790

HRSA. (2019). Area Health Resources Files. Retrieved September 11, 2019, from https://data.hrsa.gov/topics/health-workforce/ahrf

Ikegami, N. (2015). Fee-for-service payment – an evil practice that must be stamped out? *International Journal of Health Policy and Management*, Vol. 4, pp. 57–59. https://doi.org/10.15171/ijhpm.2015.26

Jacobson, M., Earle, C. C., Price, M., & Newhouse, J. P. (2010). How Medicare's Payment Cuts For Cancer Chemotherapy Drugs Changed Patterns Of Treatment. *Health Affairs*, *29*(7), 1391–1399. https://doi.org/10.1377/hlthaff.2009.0563

Johns Hopkins. (2016). Institute for Clinical and Translational Research: Truven Health MarketScan Research Databases. Retrieved September 11, 2019, from https://ictr.johnshopkins.edu/programs_resources/programs-resources/research-participant-recruitment-and-retention/truven-health-marketscan-research-databases/

Joint Commission. (2019). Measure Information Form: Perinatal Care-01. Retrieved May 11, 2020, from https://manual.jointcommission.org/releases/TJC2019A/MIF0166.html

Jürges, B. H. (2009). *Health and Old Age Health Insurance Status*. *129*, 297–307.

Kahn-Lang, A., & Lang, K. (2018). *The Promise and Pitfalls of Differences-in-Differences: Reflections on 16 and Pregnant and Other Applications*.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kamal, R., & Cox, C. (2018). *How has U.S. spending on healthcare changed over time?* Retrieved from https://www.healthsystemtracker.org/chart-collection/u-s-spending-healthcare-changed-time/?_sf_s=changed

Kessler, D., & McClellan, M. (1996). Do Doctors Practice Defensive Medicine? *The Quarterly Journal of Economics*, *111*(2), 353–390. https://doi.org/10.2307/2946682

Lakdawalla, D., & Philipson, T. (1998). Nonprofit production and competition. *NBER Working Paper No.6377*. Retrieved from https://www.nber.org/system/files/working_papers/w6377/w6377.pdf

Lakdawalla, D., & Philipson, T. (2006). The nonprofit sector and industry performance. *Journal of Public Economics*, *90*(8–9), 1681–1698. https://doi.org/10.1016/j.jpubeco.2005.11.004

Liao, J. M., Sommers, B. D., & Navathe, A. S. (2018). Medicaid's Path to Value-Based Reform. *New England Journal of Medicine*, *379*(2), 105–108. https://doi.org/10.1056/NEJMp1801942

Lungen, M., Stollenwerk, B., Messner, P., Lauterbach, K. W., & Gerber, A. (2008). Waiting times for elective treatments according to insurance status: A randomized empirical study in Germany. *International Journal for Equity in Health*, *7*(1), 1. https://doi.org/10.1186/1475-9276-7-1

Mackinnon, J. G. (2015). Wild cluster bootstrap confidence intervals. *Identification, Simulation, and Finite-Sample Inference*, *91*(1–2), 11–33. https://doi.org/10.7202/1036912ar

Mafi, J. N., Russell, K., Bortz, B. A., Dachary, M., Hazel, W. A., & Fendrick, A. M. (2017). Low-Cost, High-Volume Health Services Contribute The Most To Unnecessary Health Spending. *Health Affairs*, *36*(10), 1701–1704. https://doi.org/10.1377/hlthaff.2017.0385

Main, E., Oshiro, B., Chagolla, B., Bingham, D., Dang-Kilduff, L., Kowalewski, L., & Elimi, K. L. (2010). *Elimination of Non-medically Indicated (Elective) Deliveries Before 39 Weeks Gestational Age*.

Maratt, J. K., Kerr, E. A., Klamerus, M. L., Lohman, S. E., Froehlich, W., Bhatia, R. S., & Saini, S. D. (2019, September 15). Measures Used to Assess the Impact of Interventions to Reduce Low-Value Care: a Systematic Review. *Journal of General Internal Medicine*, Vol. 34, pp. 1857–1864. https://doi.org/10.1007/s11606-019-05069-5

McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., & Kerr, E. A. (2003). The Quality of Health Care Delivered to Adults in the United States. *New England Journal of Medicine*, *348*(26), 2635–2645. https://doi.org/10.1056/NEJMsa022615

McGuire, T. G., & Pauly, M. V. (1991). Physician response to fee changes with multiple payers. In *Journal of Health Economics* (Vol. 10). Retrieved from https://pdfs.semanticscholar.org/8975/60c8712403c298255e600ec059397260e2b2.pdf

MedPAC. (2019). *MedPAC: Report to the Congross on Medicare Payment Policy 2019*. Retrieved from www.medpac.gov.

Mendelson, A., Kondo, K., Damberg, C., Low, A., Motúapuaka, M., Freeman, M., … Kansagara, D. (2017). The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care. *Annals of Internal Medicine*, *166*(5), 341. https://doi.org/10.7326/M16-1881

Moon, J., & Shugan, S. M. (2020). Nonprofit Versus For-Profit Health Care Competition: How Service Mix Makes Nonprofit Hospitals More Profitable. *Journal of Marketing Research*, *57*(2), 193–210. https://doi.org/10.1177/0022243719901169

National Conference of State Legislatures. (2020). Maternity Length of Stay Rules. Retrieved May 11, 2020, from https://www.ncsl.org/research/health/final-maternity-length-of-stay-rules-published.aspx

National Practitioners Data Bank. (2019). The NPDB: Public Use Data File. Retrieved May 11, 2020, from https://www.npdb.hrsa.gov/resources/publicData.jsp

National Quality Forum. (2019). Quality Positioning System. Retrieved May 11, 2020, from http://www.qualityforum.org/QPS/QPSTool.aspx#qpsPageState=%7B%22TabType%22%3A1,%22TabContentType%22%3A1,%22SearchCriteriaForStandard%22%3A%7B%22TaxonomyIDs%22%3A%5B%5D,%22SelectedTypeAheadFilterOption%22%3Anull,%22Keyword%22%3A%22%22,%22PageSize%22%3A%222

Navathe, A. S., Liao, J. M., Polsky, D., Shah, Y., Huang, Q., Zhu, J., … Emanuel, E. J. (2018). Comparison Of Hospitals Participating In Medicare's Voluntary And Mandatory Orthopedic Bundle Programs. *Health Affairs*, *37*(6), 854–863. https://doi.org/10.1377/hlthaff.2017.1358

Navathe, A. S., Troxel, A. B., Liao, J. M., Nan, N., Zhu, J., Zhong, W., & Emanuel, E. J. (2017). Cost of Joint Replacement Using Bundled Payment Models. *JAMA Internal Medicine*, *177*(2), 214. https://doi.org/10.1001/jamainternmed.2016.8263

Newhouse, J P, & Marquis, M. S. (1978). The norms hypothesis and the demand for medical care. *The Journal of Human Resources*, *13 Suppl*, 159–182. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/722065

Newhouse, Joseph P. (1970). Towards a Theory of Nonprofit Institutions: An Economic of a Hospital. *American Economic Review*, *60*(1), 64–74. Retrieved from https://www.jstor.org/stable/pdf/1807855.pdf

Newhouse, Joseph P. (1970). Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital. *American Economic Review*, *60*(1), 64–74. https://doi.org/10.1126/science.151.3712.867-a

Rizzo, J. A., Zeckhauser, R. J., Kennedy, J. F., Avins, M., Borjas, G., Burgess, J., … Waldfogel, J. (2002). *Reference Incomes, Loss Aversion, and Physician Behavior*. Retrieved from https://pdfs.semanticscholar.org/1a46/ec99f7606eacd2c5db9cce9c3b55f0ec6f6a.pdf

Roodman, D., Nielsen, M. Ø., Webb, M. D., & Mackinnon, J. G. (2018). *Fast and Wild: Bootstrap Inference in Stata using boottest*.

Rosko, M., Al-Amin, M., & Tavakoli, M. (2020). Efficiency and profitability in US not-for-profit hospitals. *International Journal of Health Economics and Management*, *20*(4), 359–379. https://doi.org/10.1007/s10754-020-09284-0

Roth, J. (2019). *Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends*.

Ryan, A. M., Burgess, J. F., & Dimick, J. B. (2015). Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences. *Health Services Research*, *50*(4), 1211–1235. https://doi.org/10.1111/1475-6773.12270

Schwierz, C., Wübker, A., Wübker, A., & Kuchinke, B. A. (2011). Discrimination in waiting times by insurance type and financial soundness of German acute care hospitals. *European Journal of Health Economics*, *12*, 405–416. https://doi.org/10.1007/s10198-010-0254-2

Smith, M., Saunders, R., Stuckhardt, L., & McGinnis, J. M. (2013). A Continuously Learning Health Care System. In *Best Care at Lower Cost: The Path to Continuously Learning Healthcare in America*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK207218/?report=reader

Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Author Manuscript Health Serv Outcomes Res Methodol*, *14*(4), 166–182. https://doi.org/10.1007/s10742-014-0123-z

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8 SUPPL.8). https://doi.org/10.1016/j.jclinepi.2013.01.013

Sun, L., & Abraham, S. (2020). *Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects*.

Tai-Seale, M., McGuire, T. G., & Zhang, W. (2007). Time allocation in primary care office visits. *Health Services Research*, *42*(5), 1871–1894. https://doi.org/10.1111/j.1475-6773.2006.00689.x

Teleki, S. (2017). Birthing a Movement To Reduce Unnecessary C-Sections: An Update From California. *Health Affairs Blog*. https://doi.org/10.1377/hblog20171031.709216

The Henry J. Kaiser Family Foundation. (2018). *State Health Facts: Hospitals by Ownership Type*. Retrieved from https://www.kff.org/other/state-indicator/hospitals-by-ownership/?dataView=0&activeTab=map&currentTimeframe=0&selectedDistributions=for-profit&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

The Henry J. Kaiser Family Foundation. (2019). *Births Financed by Medicaid*. Retrieved from https://www.kff.org/medicaid/state-indicator/births-financed-by-medicaid/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

Town, R., Wholey, D. R., Kralewski, J., & Dowd, B. (2004). Assessing the Influence of Incentives on Physicians and Medical Groups. *Medical Care Research and Review*, *61*(3), 80–118. https://doi.org/10.1177/1077558704267507

Truven Health Analytics. (2013). *The Cost of Having a Baby in the United States*. Retrieved from http://www.chqpr.org/downloads/CostofHavingaBaby.pdf

U.S. Census Bureau. (2018). *American Community Survey*. Retrieved from https://www.census.gov/programs-surveys/acs

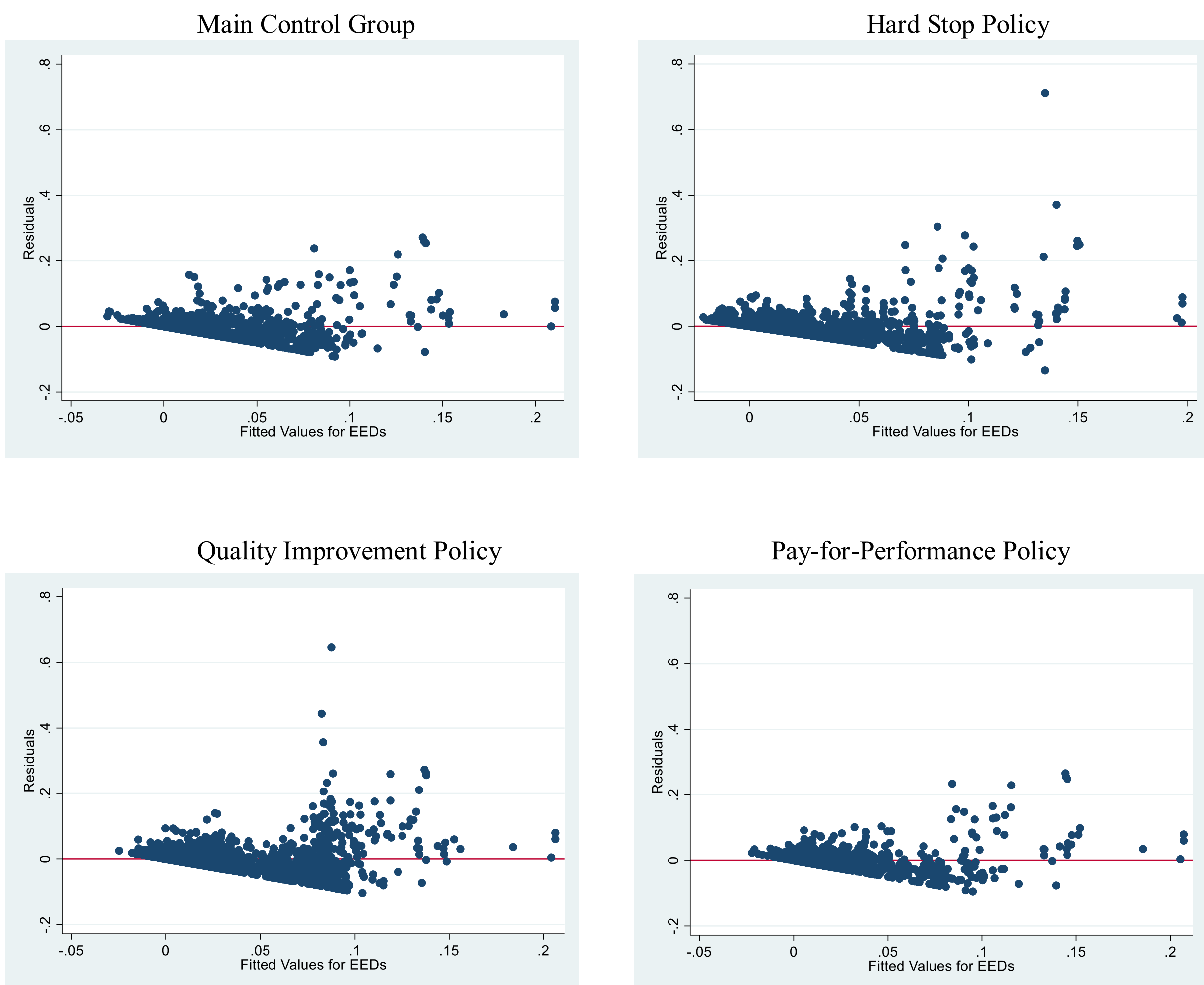
Wilson, R., & Din, A. (2018). Understanding and Enhancing the U.S. Department of Housing and Urban Development’s ZIP Code Crosswalk Files. In *Cityscape: A Journal of Policy Development and Research* • (Vol. 20).

Yip, W. C. (1998). Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors. *Journal of Health Economics*, *17*(6), 675–699. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10339248

2.10. Appendix Figures and Tables

2.10.1. Appendix Figures

Figure 2-3 Residual versus Fitted Plot Values for Early Elective Deliveries

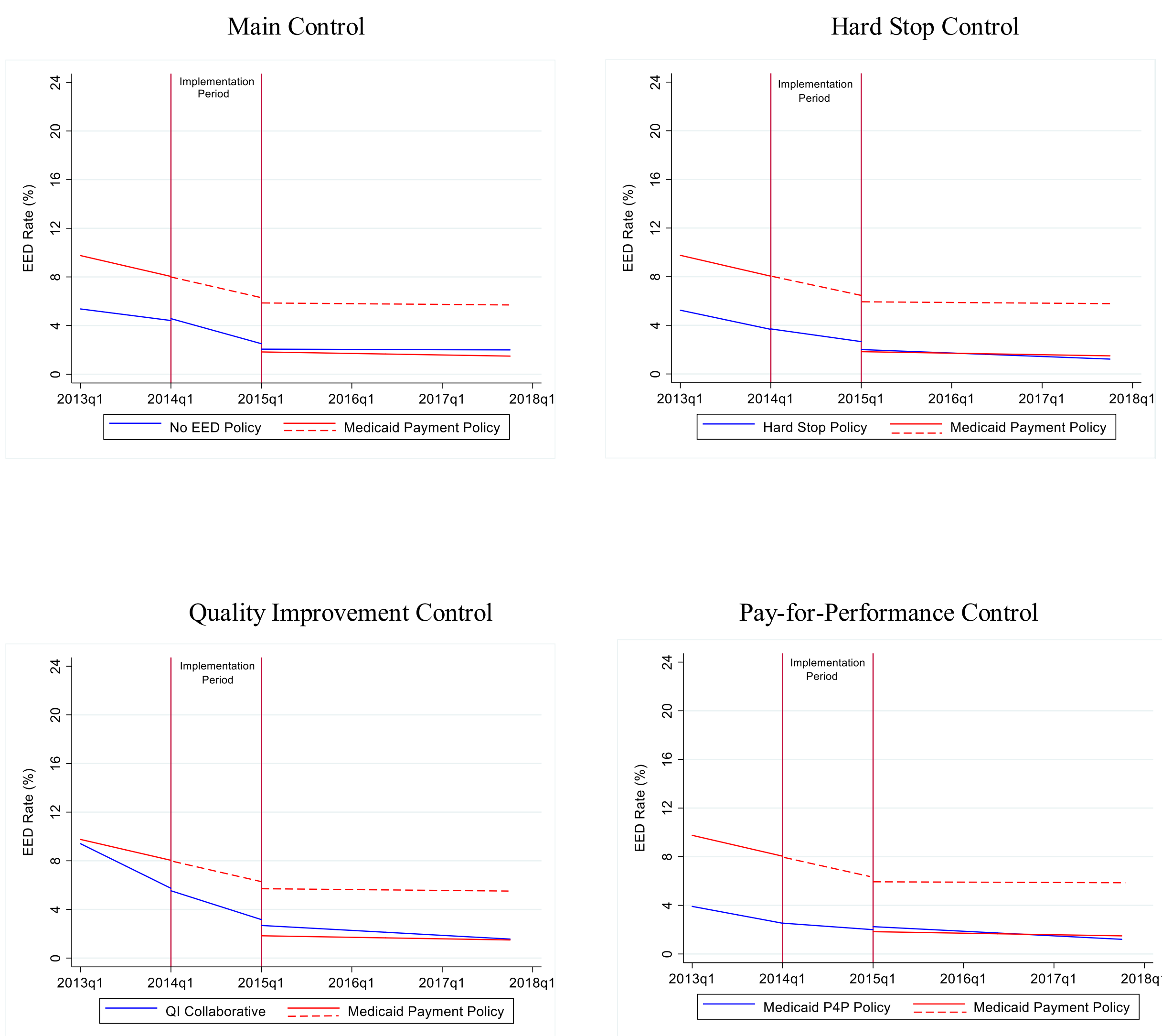


Notes: Sample estimates for EEDs are from Hospital Compare in 2013-2017. The residual versus fitted scatter plot maps the residuals from the main DD regression on the y-axis and the predicted values of the outcome from the main DD regression on the x-axis to assess whether the distribution of standard errors is similar at each value of the outcome. A scatter plot with a "random cloud" around the line y = 0 reflects homoscedasticity. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects.

Figure 2-4 Residual versus Fitted Value Plots for Low-Risk Cesarean Sections

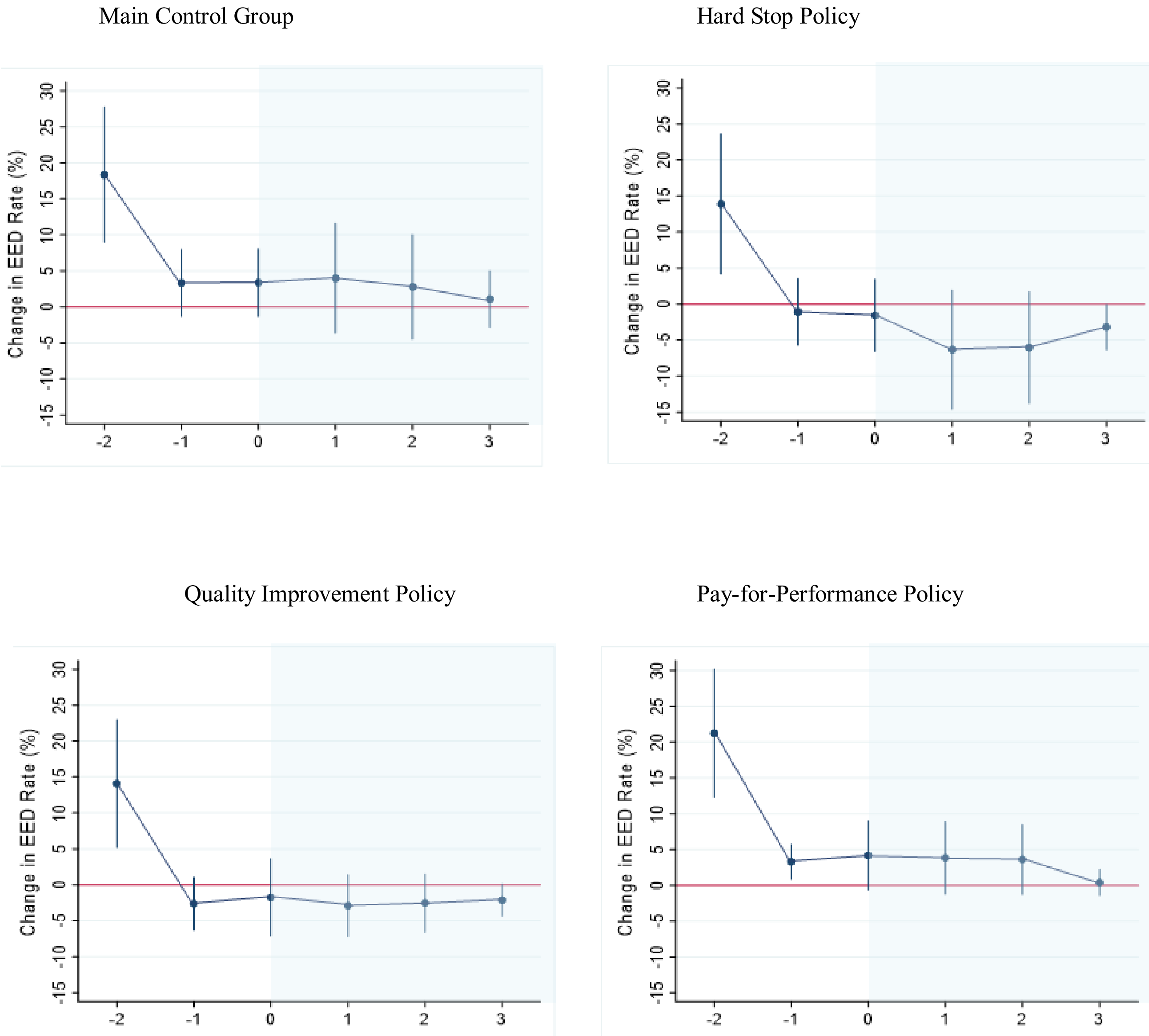
Notes: Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. The residual versus fitted scatter plot maps the residuals from the main DD regression on the y-axis and the predicted values of the outcome from the main DD regression on the x-axis to assess whether the distribution of standard errors is similar at each value of the outcome. A scatter plot with a "random cloud" around the line y = 0 reflects homoscedasticity. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects.

Figure 2-5 Simulated Effect of Choosing Wisely Guidelines on Early Elective Deliveries in Treatment and Control States



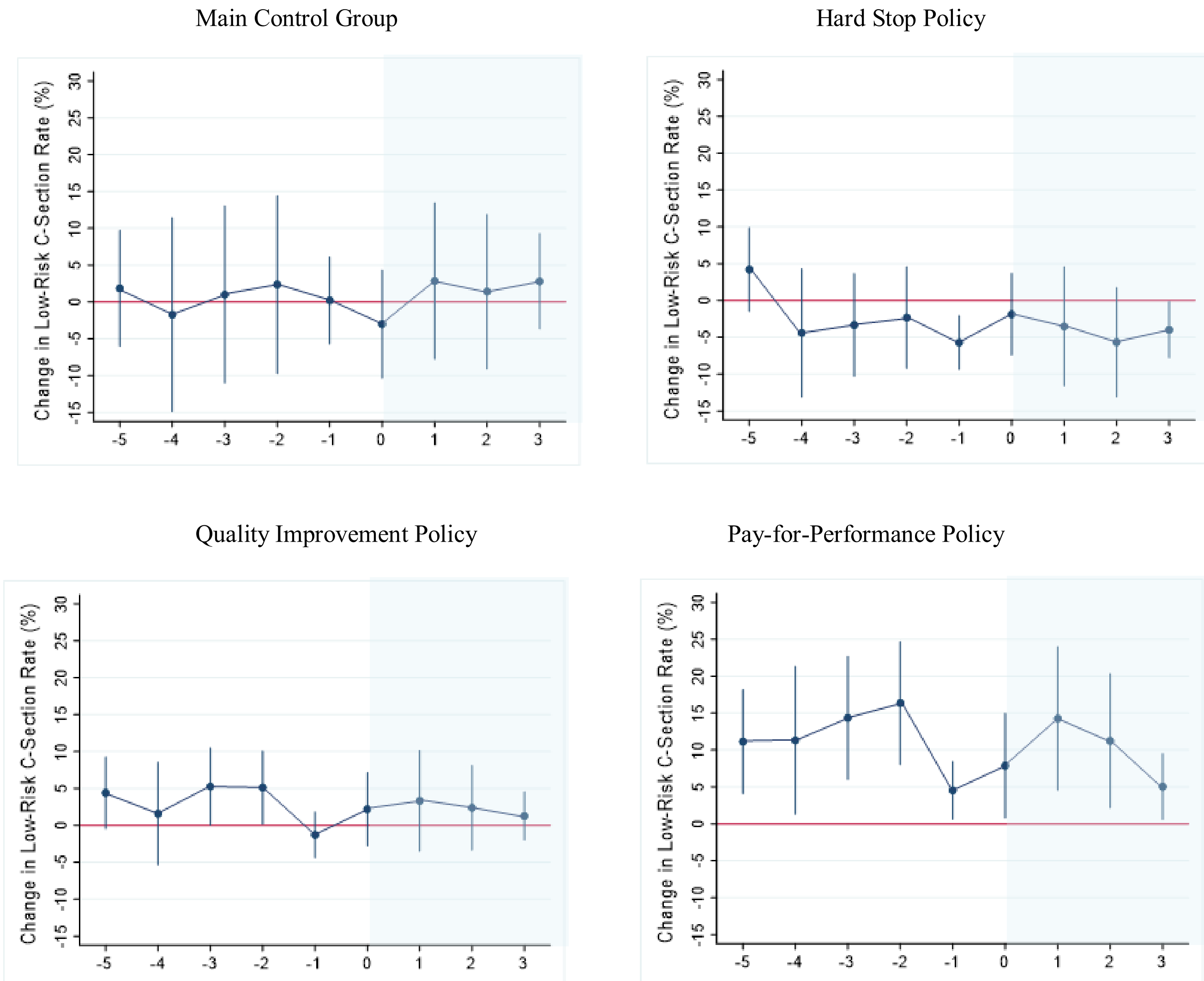
Notes: Dotted lines show the simulated effect of federal policy aimed at reducing EEDs and low-risk c-sections, including 2014 guidelines from the Choosing Wisely Campaign and the American College of Obstetricians and Gynecologists. Sample estimates for EEDs are from Hospital Compare in 2013-2017. Data points are unadjusted, quarterly means. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI).

Figure 2-6 Sensitivity Analysis - Event Study on Early Elective Deliveries



Notes: Sample estimates for EEDs are from Hospital Compare in 2013-2017. Data points are adjusted coefficients for treatment*year with 95% Confidence Intervals. Estimates overlapping 0 are not significant at the p=0.05 level. The event study design follows Callaway and Sant'Anna (2020), which measures the "group-time average treatment effect" by adding interaction terms multiplying the treatment, post-period, and year of treatment implementation. For control groups, I add a binary indicator for "never-treated." Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects.

Figure 2-7 Sensitivity Analysis - Event Study on Low-Risk Cesarean Sections



Notes: Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Data points are adjusted coefficients for treatment*year with 95% Confidence Intervals. Estimates overlapping 0 are not significant at the p=0.05 level. The event study design follows Callaway and Sant'Anna (2020), which measures the "group-time average treatment effect" by adding interaction terms multiplying the treatment, post-period, and year of treatment implementation. For control groups, I add a binary indicator for "never-treated." Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects.

2.10.2. Appendix Tables

Table 2-5 Tests for Equality in Pre-Medicaid Payment Policy Trends in Low-Value Care Outcomes

	Main	**Hard Stop**	**QI**	**P4P**
Early Elective Deliveries (2013)				
Treatment * Quarter	0.35 (-0.52, 1.23)	0.41 (-0.47, 1.31)	0.65 (-0.39, 1.70)	0.17 (-0.69, 1.09)
N	242	413	470	252
Dependent Variable Mean: % (SD)				
Treatment	9.26 (9.50)	9.01 (9.29)	9.01 (9.29)	9.01 (9.29)
Control	4.96 (4.83)	4.60 (7.91)	8.01 (8.45)	3.39 (3.20)
Low-Risk C-Sections (2010-2013)				
Treatment * Quarter	-0.06 (-0.50, 0.36)	-0.16 (-0.38, 0.06)	-0.09 (-0.31, 1.24)	0.07 (-0.25, 0.38)
N	1,026	1,746	1,986	1,051
Dependent Variable Mean: % (SD)				
Treatment	19.80 (6.67)	19.80 (6.67)	19.80 (6.67)	19.80 (6.67)
Control	18.17 (9.23)	17.59 (7.81)	18.38 (6.20)	13.76 (7.05)

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2013. Table cells include regression coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2-6 Sensitivity Analysis - Spillover Effects of Medicaid Payment Policy: Propensity Score Weighted Approach

	Main	Hard Stop	QI	P4P
Early Elective Deliveries				
Treatment * Post	-3.27 (-7.16, 0.89)	-1.82 (-5.64, 2.04)	0.59 (-3.07, 4.05)	-2.98* (-6.14, 0.28)
N	989	1,688	1,905	1,028
Low-Risk C-Sections				
Treatment * Post	0.01 (-3.15, 3.33)	-0.33 (-3.06, 2.30)	-0.57 (-3.09, 2.01)	1.10 (-3.95, 5.36)
N	1,773	3,021	3,421	1,827

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013-2017. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2-7 Covariate Balance Before and After Multiple Group Propensity Score Weights

	Pre-Medicaid Policy (2013)						**Post-Medicaid Payment Policy (2015-2017)**					
	Initial Balance			*Propensity Score Balance*			*Initial Balance*			*Propensity Score Balance*		
	Treat.	Control	SMD	Treat.	Control	SMD	Treat.	Control	SMD	Treat.	Control	SMD
Main Control												
% < HS Education	0.62	0.32	0.64	0.62	0.71	-0.18	0.65	0.27	0.82	0.65	0.89	-0.51
% >College Education	0.40	0.72	-0.67	0.40	0.31	0.18	0.45	0.75	-0.65	0.45	0.14	0.66
% Population Black	0.70	0.35	0.74	0.70	0.76	-0.12	0.71	0.34	0.81	0.71	0.85	-0.30
% Uninsured	0.65	0.28	0.78	0.65	0.70	-0.10	0.82	0.61	0.47	0.82	0.94	-0.27
% Unemployed	0.57	0.13	1.02	0.57	0.65	-0.18	0.50	0.15	0.81	0.50	0.82	-0.73
% Poverty	0.65	0.25	0.88	0.65	0.71	-0.13	0.63	0.23	0.88	0.63	0.89	-0.58
% Hospitals Non-Profit	0.41	0.55	-0.29	0.41	0.35	0.11	0.38	0.53	-0.30	0.38	0.12	0.53
% Hospitals Provide OB Services	0.49	0.44	0.09	0.49	0.77	-0.58	0.44	0.35	0.19	0.44	0.87	-0.88
Beds per 1,000	0.62	0.73	-0.22	0.62	0.79	-0.36	0.65	0.73	-0.18	0.65	0.87	-0.48
PCPs per 1,000	0.38	0.63	-0.51	0.38	0.33	0.10	0.40	0.71	-0.66	0.40	0.15	0.53
% Maternal Age 35+	0.45	0.48	-0.05	0.45	0.32	-0.27	0.47	0.49	-0.04	0.47	0.47	0.02
Avg. % Cost Sharing	0.30	0.44	-0.29	0.30	0.63	-0.67	0.50	0.43	0.13	0.50	0.55	-0.10
Hard Stop												
% < HS Education	0.62	0.52	0.21	0.62	0.62	0.00	0.65	0.53	0.25	0.65	0.65	0.01
% > College Education	0.40	0.53	-0.26	0.40	0.45	-0.10	0.45	0.52	-0.16	0.45	0.47	-0.04
% Population Black	0.70	0.47	0.48	0.70	0.68	0.04	0.71	0.47	0.50	0.71	0.71	0.01
% Uninsured	0.65	0.56	0.18	0.65	0.63	0.03	0.82	0.41	0.94	0.82	0.83	-0.01
% Unemployed	0.57	0.66	0.19	0.57	0.54	0.06	0.50	0.52	-0.04	0.50	0.44	0.12
% Poverty	0.65	0.64	0.01	0.65	0.60	0.10	0.63	0.58	0.11	0.63	0.59	0.08
% Hospitals Non-Profit	0.41	0.51	-0.22	0.41	0.29	0.23	0.38	0.48	-0.20	0.38	0.31	0.14

% Hospitals Provide OB Services	0.49	0.65	-0.35	0.49	0.50	-0.03	0.44	0.60	0.33	0.44	0.41	0.05
Beds per 1,000	0.62	0.48	0.28	0.62	0.61	0.03	0.65	0.50	0.31	0.65	0.67	-0.05
PCPs per 1,000	0.38	0.56	-0.37	0.38	0.33	0.10	0.40	0.57	-0.34	0.40	0.41	-0.03
% Maternal Age 35+	0.45	0.57	-0.23	0.45	0.44	0.02	0.47	0.55	-0.15	0.47	0.46	0.02
Avg. % Cost Sharing	0.30	0.60	-0.62	0.30	0.29	0.03	0.50	0.48	0.04	0.50	0.42	0.15
QI												
% < HS Education	0.62	0.39	0.47	0.62	0.61	0.03	0.65	0.41	0.51	0.65	0.62	0.07
% > College Education	0.40	0.56	-0.33	0.40	0.38	0.04	0.45	0.56	-0.23	0.45	0.44	0.01
% Population Black	0.70	0.70	0.00	0.70	0.72	-0.05	0.71	0.71	0.01	0.71	0.75	-0.09
% Uninsured	0.65	0.30	0.74	0.65	0.58	0.14	0.82	0.42	0.90	0.82	0.83	-0.01
% Unemployed	0.57	0.43	0.28	0.57	0.47	0.20	0.50	0.53	-0.05	0.50	0.47	0.07
% Poverty	0.65	0.41	0.49	0.65	0.54	0.22	0.63	0.41	0.46	0.63	0.59	0.08
% Hospitals Non-Profit	0.41	0.53	-0.24	0.41	0.38	0.05	0.38	0.52	-0.28	0.38	0.30	0.17
% Hospitals Provide OB Services	0.49	0.47	0.03	0.49	0.44	0.09	0.44	0.49	-0.09	0.44	0.46	-0.05
Beds per 1,000	0.62	0.57	0.11	0.62	0.58	0.08	0.65	0.58	0.14	0.65	0.62	0.06
PCPs per 1,000	0.38	0.54	-0.34	0.38	0.33	0.10	0.40	0.56	-0.31	0.40	0.37	0.06
% Maternal Age 35+	0.45	0.49	-0.08	0.45	0.48	-0.06	0.47	0.51	-0.07	0.47	0.47	0.01
Avg. % Cost Sharing	0.30	0.59	-0.60	0.30	0.39	-0.18	0.50	0.60	-0.21	0.50	0.51	-0.02
P4P												
% Less Than HS Education	0.62	0.17	1.05	0.62	0.83	-0.49	0.65	0.21	0.99	0.65	0.34	0.71
% More Than College Education	0.40	0.63	-0.48	0.40	0.16	0.49	0.45	0.56	-0.23	0.45	0.30	0.29
% Population Black	0.70	0.07	1.72	0.70	0.36	0.93	0.71	0.07	1.75	0.71	0.28	1.19
% Uninsured	0.65	0.28	0.78	0.65	0.58	0.15	0.82	0.12	1.96	0.82	0.60	0.62
% Unemployed	0.57	0.37	0.41	0.57	0.81	-0.50	0.50	0.39	0.23	0.50	0.20	0.61

% Poverty	0.65	0.16	1.15	0.65	0.53	0.29	0.63	0.16	1.11	0.63	0.57	0.13
% Hospitals Non-Profit	0.41	0.70	-0.62	0.41	0.41	-0.02	0.38	0.69	-0.64	0.38	0.57	-0.39
% Hospitals Provide OB Services	0.49	0.73	-0.52	0.49	0.60	-0.23	0.44	0.75	-0.68	0.44	0.28	0.34
Beds per 1,000	0.62	0.33	0.60	0.62	0.60	-0.04	0.65	0.30	0.74	0.65	0.56	0.19
PCPs per 1,000	0.38	0.53	-0.31	0.38	0.16	0.44	0.40	0.50	-0.30	0.40	0.46	-0.13
% Maternal Age 35+	0.45	0.44	0.02	0.45	0.43	0.04	0.47	0.50	-0.06	0.47	0.50	-0.05
Avg. % Cost Sharing	0.30	0.55	-0.51	0.30	0.23	0.16	0.50	0.67	-0.36	0.50	0.49	0.01

Notes: Propensity scores constructed using logistic regression. Covariates include binary indicators for each variable, coded as 1 if the value in a given MSA-Quarter is ≥ median and 0 if < median. Select covariates are excluded due to potential influence from the Medicaid payment policy in the post-period, including Medicaid share, commercial price difference, and OB malpractice payout. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). The estimate of interest is the Standardized Mean Difference (SMD), which provides an independent comparison between treated and control means.

Table 2-8 Sensitivity Analysis - Spillover Effects of Medicaid Payment Policy: Individual Quarterly Measures of Low-Value Care Outcomes

	Main	**Hard Stop**	**QI**	**P4P**
Early Elective Deliveries				
Treatment * Post	-2.96* (-6.41, 0.37)	-3.64*** (-6.30, -1.01)	-1.25 (-3.89, 1.40)	-4.07*** (-6.72, -1.61)
N	989	1,688	1,905	1,028
Low-Risk C-Sections				
Treatment * Post	1.30 (-1.68, 4.28)	0.43 (-1.48, 2.33)	0.49 (-1.20, 2.15)	0.95 (-1.34, 3.19)
N	1,773	3,021	3,421	1,827

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013-2017. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 2-9 Sensitivity Analysis - Spillover Effects of Medicaid Payment Policy: Alternate Treatment Groups

	Main	Hard Stop	QI	P4P
Early Elective Deliveries				
Treatment * Post				
No MO	-4.40** (-8.37, -0.59)	-4.82*** (-8.48, -1.17)	-2.32 (-5.99, 1.54)	-4.10** (-7.54, -0.55)
No MS	-2.34 (-5.91, 1.01)	-2.52* (-5.57, 0.38)	0.16 (-2.81, 3.24)	-3.34** (-6.40, -0.29)
No IN	-4.46** (-8.23, -0.67)	-5.65*** (-9.58, -1.77)	-3.07 (-7.38, 1.08)	-6.41*** (-11.25, -1.49)
No GA	-2.84 (-6.97, 0.99)	-3.32 (-7.34, 0.53)	-0.98 (-5.23, 3.05)	-3.28* (-6.67, 0.12)
Low-Risk C-Sections				
Treatment * Post				
No MO	1.15 (-1.49, 3.68)	0.16 (-1.73, 2.06)	0.22 (-1.61, 2.00)	0.74 (-1.79, 3.21)
No MS	1.53 (-1.06, 4.09)	0.44 (-1.41, 2.33)	0.11 (-1.76, 1.96)	0.83 (-1.57, 3.12)
No IN	1.22 (-1.74, 4.04)	-0.03 (-2.25, 2.05)	-0.22 (-2.31, 1.92)	0.07 (-2.86, 3.01)
No GA	-0.09 (-2.43, 2.19)	-0.56 (-2.37, 1.23)	-0.33 (-2.16, 1.53)	-0.29 (-2.62, 2.15)

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013-2017. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DD coefficients with 95% Confidence Intervals in parentheses Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 2-10 Sensitivity Analysis - Spillover Effects of Medicaid Payment Policy: Inclusion of Non-Metro State-MSAs

	Main	Hard Stop	QI	P4P
Early Elective Deliveries				
Treatment * Post	-3.54** (-6.73, -0.37)	-3.73*** (-6.67, -0.92)	-1.71 (-4.60, 1.13)	-3.92*** (-6.76, -1.12)
N	1,117	1,872	2,115	1,124
Low-Risk C-Sections				
Treatment * Post	1.05 (-1.09. 3.11)	0.24 (-1.35, 1.76)	0.09 (-1.53, 1.78)	0.59 (-1.60, 2.76)
N	1,985	3,341	3,795	1,995

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013-2017. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 2-11 Sensitivity Analysis - Spillover Effects of Medicaid Payment Policy: Exclusion of Metropolitan Statistical Areas with Attrition

	Main	Hard Stop	QI	P4P
Early Elective Deliveries				
Treatment * Post	-2.83 (-6.51, 0.88)	-4.02*** (-6.97, -1.20)	-0.64 (-4.04, 2.79)	-3.30** (-6.40, -0.26)
N	800	1,328	1,552	848
Low-Risk C-Sections				
Treatment * Post	1.67 (-1.11, 4.37)	0.87 (-0.99, 2.59)	0.95 (-0.91, 2.70)	0.40 (-1.99, 2.68)
N	1,400	2,324	2,716	1,484

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013-2017. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 2-12 Sensitivity Analysis - Spillover Effects of Medicaid Payment Policy: Event Study

	Main	**Hard Stop**	**QI**	**P4P**
Early Elective Deliveries				
Treatment * Post	-18.32***	-18.59***	-14.35***	-20.74***
	(-28.71, -7.93)	(-29.15, -8.04)	(-24.89, -3.81)	(-31.34, -10.14)
N	989	1,688	1,905	1,028
Low-Risk C-Sections				
Treatment * Post	0.86	-2.09	-0.75	-1.44
	(-3.61, 5.32)	(-5.35, 1.17)	(-3.67, 2.17)	(-6.06, 3.18)
N	1,773	3,021	3,421	1,827

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013-2017. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. The event study design follows Callaway and Sant'Anna (2020), which measures the "group-time average treatment effect" by adding interaction terms multiplying the treatment, post-period, and year of treatment implementation. For control groups, I add a binary indicator for "never-treated." Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 2-13 Sensitivity Analysis - Spillover Effects of Medicaid Payment Policy: Hospital and Individual Level Results

	Main	Hard Stop	QI	P4P
Early Elective Deliveries				
Treatment * Post	-1.76 (-5.02, 1.39)	-2.60** (-5.32, -0.31)	-1.25 (-4.48, 1.23)	-2.97** (-5.45, -0.47)
N	3,102	5,728	7,620	3,377
Low-Risk C-Sections				
Treatment * Post	0.60 (-1.62, 2.60)	-0.56 (-1.76, 0.67)	-0.20 (-1.39, 0.94)	-0.44 (-1.77, 1.22)
N	91,561	243,109	184,823	99,677

Notes: Sample estimates for EEDs from the Hospital Compare Database in 2013-2017. Sample estimates for low-risk c-sections are from Truven MarketScan claims, using data from births between 2010-2017. Table cells include DD coefficients with 95% Confidence Intervals in parentheses. Standard errors are wild cluster bootstrapped 1,000 times at the State-MSA level. Control groups are defined by policy type, including: (1) main control group with no EED policy (8 states: ID, ME, NE, NJ, RI, ND, SD, VA, WY); (2) hard stop policy group (11 states: AR, UT, DE, IA, MA, MI, MN, NC, OR, TN, OK)); (3) quality improvement collaborative (11 states: AL, AZ, CA, CT, FL, IL, KS, WV, OH, NH, and VT); and (4) Medicaid pay-for-performance group (3 states: WA, CO, WI). Covariates include all variables in Table 2-1, plus year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 2-14 Early Elective Delivery State Policy Timeline (2007-2017)

State	Policy	Implementation Date
Treatment		
Indiana	Medicaid does not cover any claim submitted for EEDs if not properly documented as medically necessary.	July 2014
Mississippi	Medicaid denies any claim submitted for EEDs if not properly documented as medically necessary.	January 2015
Missouri	Medicaid denies any claim submitted for EEDs if not properly documented as medically necessary.	October 2014
	Midwest Health Initiative: Developed and disseminated "Policy Toolkit to Support Reduction of EEDs"	2012
Georgia	Medicaid denies any claim submitted for EEDs if not properly documented as medically necessary.	January 2014
Main Control		
Idaho	N/A	N/A
Maine	N/A	N/A
Nebraska	N/A	N/A
New Jersey	N/A	N/A
Rhode Island	N/A	N/A
North Dakota	N/A	N/A
South Dakota	N/A	N/A
Virginia	N/A	N/A
Wyoming	N/A	N/A
Hard Stop Policy		
Arkansas	ARbestHealth: Program mandating all Arkansas hospitals to pledge to prevent EEDs through a hard stop policy. Hospitals voluntarily submit EED rates to the ARbestHealth Hospital Quality Team	February 2012
Utah	Maternal and Infant Health Program: Hospitals institute policies against EEDs	2009
Delaware	Delaware Healthy Mother and Infant Consortium: Mandate for hospitals to adopt guidelines to reliminate 100% of EEDs by December 2013	2011
Iowa	Iowa Hospital Engagement Network: Urged participating hospitals to pledge to reduce EEDs.	2013
Massachusetts	Massachusetts Perinatal Quality Collaborative: Urged hospitals to adopt voluntary hard stop policies.	May 2011
Michigan	Michigan Department of Community Health: implemented hard stop policy and required all Medicaid-enrolled birthing hospitals to utilize EED evidence-based guidelines	Jan 2013
	Michigan Health and Hospital Association Keystone Obstetric Collaborative: Voluntary initiative for hospitals the prohibited elective c-sections and inductions before 39 weeks gestation.	2009
Minnesota	Evidence-Based Childbirth Program: Law required hospitals to implement policies to minimize EEDs	January 2012
	Blended reimbursement rate for c-sections and vaginal deliveries	October 2009

North Carolina	Pregnancy Medical Home Program: Overarching goal to improve birth outcomes and reduce costs. To qualify for participation, hospitals must adopt hard stop policy to eliminate EEDs.	March 2011
Oregon	Oregon Perinatal Collaborative and March of Dimes 39 Weeks campaign: Urged hospitals to enact hard stop policy to eliminate EEDs	February 2012
Tennessee	Tennessee Healthy Babies are Worth the Wait: Requested that all hospital CEOs in the state sign a pledge to adopt hard stop policies and submit data on hospital EED rate	2013
Oklahoma	Every Week Counts Collaborative: Recruited hospitals for voluntary hard stop program to eliminate EEDs.	April 2011
Quality Improvement Collaborative		
Alabama	Alabama Perinatal Excellence Collaborative: Created and disseminated guidelines for scheduling deliveries before 39 weeks gestation to hospitals in the state	January 2012
Arizona	Arizona Perinatal Trust: Integrates voluntary certification of guideline adherence, perinatal education, and perinatal data analysis to improve maternal and neonatal outcomes and quality	January 2010
California	The California Maternal Quality Care Collaborative: Developed and disseminated toolkit for preventing statewide EEDs	2010
	Patient Safety First: Voluntary collaborative to reduce EED rate below 5% by 2012	January 2010
	California Hospital Engagement Network: Initiative to reduce EED rate to <3% in the state	March 2012
Connecticut	Participant of March of Dimes Perinatal Quality Improvement Initiative: Awareness campaign for obstetric providers on risks of EEDs. Tasked with integrating CMQCC Toolkit into hospitals.	2011
Florida	Florida Perinatal Quality Collaborative: Educate providers on EED risks in collaboration with March of Dimes	June 2010
Illinois	Illinois Perinatal Quality Collaborative: Quality improvement obstetric initiative focused on reducing EEDs	2012
	Midwest Health Initiative: Developed and disseminated "Policy Toolkit to Support Reduction of EEDs"	2012
	Midwest Business Group on Health: Collaborative between National Business Coalition on Health, Quality Quest for Health, the State of Illinois, and March of Dimes to prevent EEDs and improve maternal quality and outcomes	2011
Kansas	Kansas Perinatal Quality Collaborative: Quality improvement initiative aimed at eliminating EEDs	September 2012
	Kansas Healthcare Collaborative and Hospital Engagement Network: Set goal to reduce EED rate to <3% by end of 2013	July 2012
West Virginia	West Virginia Perinatal Partnership: Initiated quality improvement program to reduce EEDs. Participation consisted of 14 hospitals, representing 70% of births in the state).	2009
Ohio	Ohio Perinatal Quality Collaborative: Initiated the 39 Weeks Delivery Charter Project, which made efforts to reduce unnecessary EEDs.	2008
New Hampshire	Northern New England Perinatal Quality Improvement Network: A voluntary consortium of healthcare organizations committed to improving care for women and children. Offers education programs, best practice guidelines, benchmarking quality rates, and team-based approach to reducing poor outcomes.	2007
Vermont	Northern New England Perinatal Quality Improvement Network: A voluntary consortium of healthcare organizations committed to improving care for women and children. Offers education programs, best practice guidelines, benchmarking quality rates, and team-based approach to reducing poor outcomes.	2007

Medicaid Pay-for-Performance Program		
Washington	Safety Net Assessment Act: gave hospitals a 1% increase in their Medicaid reimbursement for reducing annual EEDs	April 2010
	Blended reimbursement rate for uncomplicated c-sections and vaginal deliveries	April 2009
	Washington State Perinatal Collaborative: Encouraged hospitals to sign pledge to reduce EEDs	November 2010
Wisconsin	Obstetric Medical Home: Pays $1,000 bonus for each Medicaid patient that attends at least ten prenatal visits and a postpartum visit within 60 days of birth. Additional $1,000 bonus per positive birth outcome.	Piloted 2011. Enacted January 2014.
	Blended reimbursement rate for c-sections and vaginal deliveries.	January 2010
Colorado	Partnership for Patients: A quality improvement program, led by the Colorado Hospital Association, aimed at reducing statewide EED rates.	April 2011
	Hospital Quality Incentive Payment Program: Offers volume-adjusted payments based on Medicaid discharges and quality achievement on EED performance.	2011
Excluded		
Louisiana	Commercial and Medicaid insurers do not reimburse any claim submitted for EEDs if not properly documented as medically necessary.	September 2014
	Excluded due to multi-insurer effort; would not measure direct spillover.	
Maryland	Maryland Perinatal System Standards: Maryland Department of Health and Mental Hygiene develop and disseminate voluntary standards and hospitals participate in hard stop policy to eliminate EEDs	July 2012
	Excluded due to Hospital Global Budget; potential contamination of policy effect.	
Montana	Medicaid denies any claim submitted for EEDs if not properly documented as medically necessary.	October 2014
	No Montana hospitals reported EED rates.	
New Mexico	Medicaid denies any claim submitted for EEDs if not properly documented as medically necessary.	January 2014
	Excluded due to Medicaid for maternity care based on one global budget; potential contamination of policy effect.	
	Medicaid offers blended reimbursement rate for c-sections and vaginal deliveries	April 2011
Nevada	Medicaid denies any claim submitted for EEDs if not properly documented as medically necessary.	June 2012
	Excluded due to lack of availability of pre-policy data (no pre-policy data available prior to 2013).	
New York	Medicaid reduced payments for EEDs by 10% unless documented as medically indicated	July 2013
	Excluded due to lack of availability of pre-policy data (no pre-policy data available prior to 2013).	
	Medicaid Redesign Team Reforms: quality improvement collaborative aimed at lowering statewide Medicaid spending. One initiative directed towards EEDs.	January 2011
South Carolina	Medicaid and BlueCross BlueShield deny payment for non-medically necessary EEDs.	January 2013
	South Carolina Birth Outcomes Initiative: Encouraged hospitals to adopt hard stop policy pledge to reduce EEDs.	March 2011
	Excluded due to lack of availability of pre-policy data (no pre-policy data available prior to 2013).	
Texas	Texas House Bill 1983 required all hospitals to implement practices to reduce EEDs	September 2011
	Medicaid denies any claim submitted for EEDs if not properly documented as medically necessary.	June 2011
	Excluded due to lack of availability of pre-policy data (no pre-policy data available prior to 2013).	

Kentucky	Medicaid EEDs require prior authorization. **Excluded due to potential for policy contamination; policy implementation occurs after 2015.**	September 2017
Pennsylvania	Pennsylvania Hospital Engagement Networks Obstetric Adverse Events Collaborative: Used peer comparisons and quality reporting of EEDs to discourage their provision. **Excluded due to non-parallel pre-policy trends in quality improvement comparison group; suggests unobservable differences in PA could drive policy adoption and effects.**	May 2013

3. Impacts of Mandatory Bundled Payments on Racial Disparities

3.1. Introduction

Decades of research has centered on racial disparities in healthcare and potential solutions, yet inequities remain pervasive (Kumar, Mahmoudi, & Rivera-Hernandez, 2019; Wasserman et al., 2019). However, most existing work does not account for residential segregation at the metropolitan level, which generates significant variation in quality and access to care across settings (LaVeist, Pollack, Thorpe, Fesahazion, & Gaskin, 2011). Indeed, it is clear that the interaction of social determinants and structural inequities, economic and physical environments, and implicit biases among clinicians and healthcare organizations, create barriers that hinder timely receipt of effective clinical care, and often lead to poorer outcomes among minority patients (Damberg et al., 2015). Since these mechanisms are overlapping, focusing on effects of individual race without accounting for place-based racial composition hinders the ability to detect how the race and socioeconomic drivers combined produce disparities in health status and its indicators (LaVeist et al., 2011). In turn, healthcare policies targeting racial gaps in care delivery may be limited due to the confounding effects between geographic racial composition and socioeconomic status.

One policy approach to improving value in healthcare is moving payment systems away from fee-for-service (FFS) towards reimbursement mechanisms that align spending with quality. Non-FFS payment reforms have potential to address racial disparities by providing physicians with a financial incentive to change their behavior towards minority populations, thereby increasing healthcare quality and narrowing gaps in care (Damberg et al., 2015). Efforts to design such incentives have predominantly focused on improving quality for the “average” patient, rather than vulnerable populations such as racial minorities. Inattention to heterogeneity in effects across populations can lead to an inadequate understanding of the impacts of payment reforms. This may hamper the potential to leverage financial incentives as a tool for alleviating disparities. Evaluating the impact of value-based payment reforms on place-based racial disparities can highlight current policy gaps in physician payment, and specifically, whether health reforms aimed at improving quality of care for the average patient narrow, or inadvertently widen, racial inequities.

Evidence exploring the effects of value-based incentives on vulnerable populations is mixed. Some studies show that value-based payment has no effect on racial and socioeconomic disparities with respect to quality of care (Hsu et al., 2020; Stone, 2020). Other studies have found that providers who care for a disproportionate share of disadvantaged patients, including those who are dually eligible for Medicare and Medicaid, low-income, and racial minorities, tend to exhibit smaller gains in quality improvement, in turn widening disparities (Chaiyachati, Qi, & Werner, 2018; Colla et al., 2012; Figueroa, Zheng, John Orav, Epstein, & Jha, 2018; McWilliams, Chernew, & Landon, 2017; Song, Rose, Chernew, & Safran, 2017). This result tends to be most prevalent when payment reform incorporates downside risk (Chaiyachati et al., 2018; Figueroa et al., 2018; Gaskin, Zare, Vazin, Love, & Steinwachs, 2018; Gilman et al., 2014; Gilman, Hockenberry, et al., 2015; Gu et al., 2014; Joynt & Jha, 2011, 2013; Joynt Maddox, 2018; Joynt Maddox, Reidhead, Qi, & Nerenz, 2019), which is problematic because it has potential to redistribute payment away from providers with the highest resource needs (Damberg et al., 2015).

Thus, value-based payment has potential to initiate a cycle of worsening disparities for socially and clinically vulnerable groups if the benefits of value-based care do not extend to these populations (Chaiyachati et al., 2018). Unintended consequences, including avoidance of vulnerable populations, are also possible, if physicians assume that treating these patients increases the likelihood of financial penalty (Hsiang et al., 2019; Lee, Polsky, Fitzsimmons, & Werner, 2020; Werbeck, Wübker, & Ziebarth, 2020; Werner, Kanter, & Polsky, 2019; Yasaitis, Pajerowski, Polsky, & Werner, 2016). A large literature explores the impacts of non-FFS payment reforms on racial and socioeconomic disparities separately, but only a small number of studies assess their effect on income and race-based inequities combined. Addressing this interaction can help to determine potential ways to leverage physician reimbursement policy to reduce variation in healthcare quality across vulnerable populations.

Exploring potential solutions for racial disparities is particularly important in perinatal care. While many high income countries have seen a decline in maternal deaths in recent years, the U.S. rate continues to climb (Rana, Alam, & Gow, 2018; Shaw et al., 2016). This issue is especially pervasive among Black mothers, who experience 3.7 times more deaths from pregnancy-related complications and more than twice the rate of severe maternal morbidity compared to White mothers (Howell, Egorova, Balbierz, Zeitlin, & Hebert, 2016; Kozhimannil et al., 2017; Min, Ehrenthal, & Strobino, 2015; Oribhabor, Nelson, Buchanan-Peart, & Cancarevic, 2020). Such deaths are often preventable, and racial disparities persist across income and education levels. Effective clinical care is critical for optimizing perinatal outcomes, yet for Black mothers, the care they receive often falls short of their medical needs. Black women are 20% less likely to receive adequate prenatal and postpartum care, and 5% more likely to receive low-value services like unnecessary cesarean sections compared to other racial groups (Getahun et al., 2009; Min et al., 2015). These treatment patterns suggest that physician behavior is a potential driver, increasing the likelihood for financial incentives to address these inequities.

I build on prior literature by assessing the impact of a state payment reform on place-based racial disparities in perinatal care. Using the 2010 to 2016 Truven MarketScan Commercial Claims database, I study whether a multi-payer mandatory bundled payment program in Arkansas affected areal-level racial disparities in perinatal treatment among three types of services: (1) low-risk c-sections, (2) appropriate prenatal screenings, and (3) timely postpartum visits. I use a difference-in-differences model to compare changes in perinatal care in Arkansas versus a group of control states created via multiple propensity score weights, in Metropolitan Statistical Areas (MSAs) with a high proportion of Black patients and MSAs with a high proportion of White patients. Implementation of the Arkansas episode-based payment program provides two broad results: (1) in the short-term, MSAs with a high proportion of Black residents exhibit worsening quality relative to MSAs with a high proportion of White residents, and (2) in the long-term, there are no significant effects on place-based racial disparities. The role of financial incentives on disparities has only been assessed in clinical settings that rely on a binary indicator for receiving treatment; however, maternal care allows me to test whether areas with a high proportion of Black patients receive inferior treatment across a range of treatment options. In addition, by focusing on place-based, rather than individual disparities, I account for the joint effects of residential segregation based race and income, and in turn, barriers to accessing care across geographic environments (LaVeist et al., 2011). This study aims to refine the understanding of the role of physician incentives on disparities, which is a critically important question given current inequalities in maternal care and the potential for payment policies to help close or widen this gap.

The paper proceeds as follows. Section 3.2 provides background on the Arkansas episode-based payment program. Section 3.3 describes the methods, including the analytic dataset, key variables, and empirical strategy. Section 3.4 discusses the main results, and Section 3.5 concludes.

3.2. Background

3.2.1. Arkansas Payment Improvement Initiative in Perinatal Care

In 2013, Arkansas introduced the Payment Improvement Initiative (APII), a mandatory, multi-payer bundled payment program. Its goal was to reduce complication-related perinatal spending and improve the patient experience of care through better quality, access, and reliability. By offering a single, risk-adjusted case rate for the entire episode of childbirth, rather than reimbursing providers separately for each individual service, the program introduced supply-side cost sharing for the delivering obstetrician, who is responsible for any unnecessary expenditures (C. Carroll et al., 2018).

A perinatal episode is triggered by a live birth, and spans forty weeks before through sixty days after the birth. The delivering obstetrician, called the Principal Accountable Provider (PAP), is held financially accountable for total risk-adjusted episode spending. A physician must have five or more eligible cases per year to qualify as a PAP. Risk-adjustment strives to alleviate the financial risk that PAPs face for complex patients. The algorithm is determined on an annual basis by commercial insurers on the basis of documented patient comorbidities, but the methodology is not publicly available. Extremely high-risk episodes are also excluded from APII. Instead, these cases are paid on a FFS basis. Episode exclusion is based on co-morbidities (e.g. severe preeclampsia, end stage renal disease, cystic fibrosis) and complications (e.g. non-live birth, pulmonary embolism, puerperal infection) documented within one year of the delivery (Arkansas Blue Cross Blue Shield, 2014). One concern is that PAPs may "upcode" in an effort to increase allowable spending targets and/or to exclude costly episodes (Dafny, 2005). To date, no empirical work has found evidence of upcoding in either capacity. Carroll et al. (2018) found that results were robust to inclusion of high-cost outlier episodes, and probability of upcoding did not differ significantly in treatment versus controls. These findings suggest that gaming on the basis of episode exclusion was limited (C. Carroll et al., 2018).

APII reimbursement is FFS with reconciliation. It employs a two-sided risk model that hinges upon whether a PAP's mean risk-adjusted spending is acceptable, unacceptable, or commendable, relative to a set spending threshold determined from historical statewide perinatal costs. PAPs with acceptable spending are paid FFS without any adjustments; PAPs with unacceptable spending receive a financial penalty equal to 50% of the excess spending beyond the acceptable threshold; and PAPs with commendable spending are provided a financial bonus equal to 50% of the savings. Insurers send quarterly reports to PAPs, enabling them to track risk-adjusted spending and quality. Eligibility for gainsharing is tied to achieving a minimum quality score of 80% or higher on select outpatient quality measures, including prenatal screening rates for Group B. Streptococcus (Group B. Strep), Chlamydia, and Human Immunodeficiency Virus (HIV). Otherwise, quality scores do not factor into reimbursement; however, PAPs are required to submit rates for additional prenatal screening measures (Hepatitis B, Gestational Diabetes, and Asymptomatic Bacturiuria (A. Bact.) and total c-section rate, to be included in the quarterly program reports (Arkansas Blue Cross Blue Shield, 2014). Historically, spending rates were mostly considered acceptable or commendable; in the first year of the program, only 9% of PAPs were penalized, while 69% were given a bonus

(ACHI, 2015). The program methodology was developed jointly between Medicaid and Arkansas Blue Cross Blue Shield (BCBS), the two main childbirth payers in the state. Although all private payers did not adopt the program, APII covered an estimated 70% of statewide births and 80% of births in the commercial sector. This is likely due to a consolidated private insurance market, with Arkansas BCBS possessing a 73% market share. In addition to BCBS, the third largest insurer in the state, QualChoice Arkansas, employed APII (C. Carroll et al., 2018).

Policy implementation occurred in two phases. In 2013, the program was partially launched, predominantly for Medicaid beneficiaries. At that time, Medicaid covered 67% of births in the state. In 2014, Arkansas introduced a unique Medicaid expansion to individuals with incomes below 138% of the federal poverty line (FPL) via a “private option,” which gave beneficiaries premium assistance to purchase private plans on the exchange (Allison, 2013). There is minimal evidence that Medicaid expansion affected uninsurance among pregnant women, since many were already covered by Medicaid (C. Carroll et al., 2018). However, enrollment among pregnant women shifted from Medicaid to private exchange plans, resulting in a lower share of births covered by Medicaid after the private option was introduced (down to 50%). Select insurers in the commercial sector predominantly rolled out APII in 2014 by introducing it to fully-insured and self-funded groups (Allison, 2013). In this paper, I consider the program to be partially implemented in 2013, and fully implemented starting in 2014. This approach is consistent with prior literature, and enables me to measure whether effects of the policy and its impact on racial disparities varied across each stage of implementation (C. Carroll et al., 2018). Using this approach, I am able to capture whether there were spillovers from Medicaid to the commercial market in 2013, compared to direct effects in the commercial sector alone after 2014.

To date, only one study has evaluated the impact of the APII. Carroll et al. (2018) examine how the bundled payment incentive affected quality of care across the entire population, rather than exploring heterogeneity by MSA-level racial composition. The authors find no significant changes in prenatal, intrapartum, or postpartum quality in the average population. However, more work is needed to highlight potential differences in magnitude and/or direction of effects in areas with varying populations (C. Carroll et al., 2018).

3.2.2. Effects of Payment Reform on Racial and Socioeconomic Disparities

Evidence exploring the effects of value-based incentives on vulnerable populations is mixed. With respect to income-based disparities, some studies have found that providers who care for a disproportionate share of disadvantaged patients tend to exhibit smaller gains in quality improvement. For example, Song et al. (2017) and Colla et al. (2012) demonstrated that bundled payments led to a smaller decline in unplanned readmissions in geographic areas with lower socioeconomic status and a higher proportion of patients dually enrolled in both Medicare and Medicaid, respectively (Colla et al., 2012; Song et al., 2017). In turn, providers that serve the highest quintile of dual beneficiaries were 18% more likely to receive financial penalties in the presence of downside risk (Gilman et al., 2014; Gu et al., 2014; Joynt & Jha, 2011, 2013; Joynt Maddox, 2018). Consistent with this result, McWilliams et al. (2017) also found that early savings in the Medicare Shared Savings Program were almost entirely concentrated among low-risk patients (McWilliams et al., 2017). In aggregate, this may redistribute payment away from providers with the highest resource needs (Damberg et al., 2015).

An unintended consequence of this result is "cream skimming," or the notion that physicians systematically choose patients based on characteristics other than their need for care, to bolster profits or reputation (Werbeck et al., 2020). In general, studies show that a growing number of physicians are less likely to accept low-income patients. Decker (2013) identified that approximately one-third of primary care physicians did not accept new Medicaid patients in 2011 and 2012, using the National Ambulatory Medical Care Survey Electronic Medical Records Supplement (Decker, 2013). A meta-analysis assessing 34 studies also found that Medicaid patients had a 1.6 and 3.3-fold reduced likelihood of successfully scheduling a primary care and specialty visit, respectively, compared to privately insured patients (Hsiang et al., 2019). This pattern persisted across settings and medical conditions. However, no consistent evidence has shown that the practice of “cream skimming” takes place in the context of value-based payment reforms. Werner et al. (2019) found that physicians caring for a higher proportion of socially and clinically vulnerable patients (including dual eligible, racial minority, and low-income populations) were no less likely to participate in accountable care organizations (ACOs) than physicians groups that did not (Werner et al., 2019). Building on this work, Lee et al. (2020) established that after joining an ACO, physicians did not alter the proportion of vulnerable patients that they treated, including racial minorities and low socioeconomic status patients (Lee et al., 2020). However, in direct contrast to this finding, Yasaitis et al. (2016) found that physicians who practiced in areas where a higher percentage of the population was Black, living in poverty, uninsured, disabled, or had less than a high school education, were less likely to join commercial and Medicare ACOs (Yasaitis et al., 2016). It is important to determine whether payment reforms prompt unintended consequences for socially and clinically vulnerable groups, as health disparities may worsen if the benefits of value-based care do not extend to these populations.

With respect to whether financial incentives affect racial disparities, results are mixed. Several studies have found no association between federal value-based incentive programs and changes in racial disparities for quality outcomes like healthcare-associated infection and rehospitalization rates (Hsu et al., 2020; Stone, 2020). Studies on the Hospital Readmission Reduction Program (HRRP) suggest that financial incentives may exacerbate racial disparities for medical conditions such as heart attack, heart failure, and pneumonia. In the initial implementation phase of HRRP, when only financial bonuses had been integrated, readmission rates for Black patients in safety net and non-safety net hospitals declined more rapidly, suggesting a small narrowing of racial disparities. However, once financial penalties were imposed, disparities in readmission rates worsened in safety net hospitals (Chaiyachati et al., 2018; Figueroa et al., 2018; Gaskin et al., 2018; Gilman, Hockenberry, et al., 2015; Joynt Maddox, Reidhead, Qi, et al., 2019). This line of work suggests that, at best, value-based payments neither improve, nor exacerbate, racial inequities. Since results varied across incentive structures (e.g. financial bonuses versus penalties) and clinical settings, examining the impact of value-based reimbursement with mandatory participation on racial disparities is integral to understanding whether results can be generalized to future incentive design.

While there is a large literature on the effects of value-based payment reforms on racial and socioeconomic disparities separately, few studies explore the joint impacts of financial incentives on income and race-based inequities. Addressing the interaction between socioeconomic and racial disparities is crucial to understanding the place-based context for observed treatment patterns, and potential ways to leverage physician reimbursement policy to reduce variation in healthcare quality across these populations.

3.3. Methods

3.3.1. Data

3.3.1.1. Truven MarketScan Commercial Claims

The primary data for this analysis is the Truven MarketScan Commercial Claims database from 2010 to 2016, which links paid claims and encounter data with detailed patient information across sites and types of providers over time. Although the database is a convenience sample of enrollees in commercial health plans and large self-insured firms that opt to provide their data, the MarketScan data includes proprietary commercial claims (employer and health plan) from over 36 million patient hospital discharges (Johns Hopkins, 2016). These data are collected across broad geographic areas to represent treatment patterns and costs in the U.S.

The MarketScan data are advantageous for this analysis because they use consistent enrollee identifiers over time, enabling me to track patients across the full episode of care. One limitation is that a major insurer dropped out of the MarketScan data in 2015. To avoid differential selection into the database over time, I limit the sample to the employer population, which remains stable over the study period. Relying on the employer population also aims to ensure that the sample is minimally affected by changes in the insurance policy landscape, such as Medicaid expansion or the private option in Arkansas. Another limitation is the lack of unique identifiers for payers and providers. This makes it difficult to ascertain whether the insurer was a participant in APII. The consolidated market in Arkansas increases the likelihood that most of the sample consists of covered episodes in the commercial sector. It also limits the ability to make inferences about individual physician or hospital behavior; instead, I focus on aggregate behavior among providers, and the extent to which APII impacts racial disparities at the system level.

3.3.1.2. U.S. Census American Community Survey

I measure MSA-level racial composition from the U.S. Census Bureau's American Community Survey (ACS), a dataset with publicly available measures on demographic and employment characteristics for all counties. Specifically, I calculate the percent of the population that is Black within an MSA from 2010 to 2012, and assign an MSA as "Black" if its share of Black residents is above 12.59% (the national mean during the survey period) and "White" if its share of Black residents is below this threshold (White residents are the majority). Given that the MarketScan data does not include a unique identifier for each non-metro area, the sample is restricted to micro and metropolitan areas, to ensure proper linkage of the MSA-level race variable.

There is high variation in the share of the Black population across MSAs. It is likely that Black MSAs include a large proportion of White residents, prompting concerns that treatment patterns are not driven explicitly by race. In the sample, MSAs characterized as Black have a mean Black population of 25%, while MSAs characterized as White only have a mean Black population of 4%, suggesting a wide enough gap in racial composition for differences in utilization to be a function of metropolitan area racial composition. This demonstrates that use of the MSA-level race statistic is likely to capture a significant proportion of Black patients.

3.3.2. Sample

The analytic sample consists of 158,858 perinatal episodes between 2010 and 2016. This includes 2,031 episodes in the treatment group (1,064 in Black MSAs and 967 in White MSAs), and 156,827 episodes in the control group (81,249 in Black MSAs and 75,578 in White MSAs). I use the MarketScan database to construct perinatal episodes using the methodology outlined in the Arkansas BCBS Perinatal Algorithm. First, I identify all live births between 2010 and 2016 using the relevant Diagnosis Related Group (DRG) codes, and pull all inpatient and outpatient claims in the period 40 weeks before the delivery through sixty days afterwards. I then collapse the data to the episode-level, using the date of birth for assignment to the pre- and post- periods.

The sample includes low-risk births in the MarketScan Database, as defined by the Agency for Healthcare Quality and Research (AHRQ). This population is limited to mothers with uncomplicated births (e.g. no abnormal presentation, preterm delivery, fetal death, multiple gestation, or breech procedure) who have never had a prior c-section. I exclude cases that are exempt from APII due to patient co-morbidities (e.g. sickle cell anemia, end stage renal disease, severe preeclampsia). Focusing on low-risk episodes that are reimbursed through the bundled payment program aims to confirm that all physicians in the treatment group face the same financial incentive. From a disparities standpoint, this inclusion criteria ensures that all mothers in the sample, to the best of my knowledge, should receive a consistent mix of clinical services. In other words, given that each mother in the sample has an uncomplicated delivery and lacks chronic conditions, a c-section would be considered an inappropriate treatment choice, based on the clinical information and patient risk available, regardless of race (Agency for Healthcare Quality and Research, 2016).

3.3.3. Outcome Variables

I construct eight outcome measures for each episode, all of which are adapted from the Maternity Care Performance Measure Set developed jointly by American College of Obstetricians and Gynecologists and the National Committee for Quality Assurance. Outcomes consist of the following prenatal measures: whether a patient received three screenings linked to gainsharing (Group B Strep, Chlamydia, and HIV) and three screenings not linked to gainsharing (Hepatitis B, Gestational Diabetes, and A. Bact.). PAPs must report each of the prenatal screening rates to track their performance, regardless of whether it is tied to reimbursement. There is one outcome in the postpartum period: whether the mother received any follow-up care within eight weeks of the delivery. All prenatal and postpartum outcomes are considered high-value components of the perinatal episode, which means that a higher rate is considered better.

The last outcome is the low-risk c-section rate, which is evaluated during the intrapartum period. This measure identifies whether the mother received an inappropriate c-section (e.g. if she is nulliparous, with a full-term, singleton birth in the vertex position). This measure is derived from methodology developed by AHRQ, using Inpatient Quality Indicator (IQI) 33, which is endorsed by the National Quality Forum as a consensus standard for hospital care (Agency for Healthcare Quality and Research, 2016). Unlike the other outcomes, low-risk c-sections capture a dimension of low-value care, meaning that a lower rate is considered better.

3.3.4. Covariates

Covariates include the following maternal characteristics: insurance type (HMO, PPO, POS, and high-deductible health plan), age category (less than 25, 25-29, 30-34, and 35+), whether the hospital length of stay was over four days (the number of days typically covered by the insurer), and cost sharing quartile bins (National Conference of State Legislatures, 2020). I also include a series of MSA-level controls, which account for healthcare, demographic, and economic variables that may influence physician practice patterns. Healthcare factors consist of hospital characteristics (percent of hospitals that are non-profit, percent of hospitals that provide obstetric services, percent teaching hospitals, number of federally qualified health centers (FQHCs), beds per 1,000, and percent of patients that are insured by Medicaid), practitioner information (primary care practitioners per 1,000), and malpractice risk quartile bins, defined as the average obstetric-related malpractice payout. Demographic characteristics refer to the percent of the population with less than a high school education and percent of the population with more than a college education. Finally, economic characteristics include the percent of the population that is uninsured, the percent unemployed, and the percent with income below the FPL. Covariates were selected to be consistent with Carroll et al., (2018).

Covariates were derived from several data sources, including the ACS, the Health Resources and Services Administration's Area Health Resource File (AHRF), the American Hospital Association (AHA) Annual Survey, and the National Practitioners Data Bank (NPDB) (AHA, 2018; HRSA, 2019; NPDB, 2019; U.S. Census Bureau, 2018).

3.3.5. MSA-Level Race Variable

I rely on MSA-level racial composition, rather than individual patient race, to capture differences in the effects of APII according to variation in the percent of the population that is Black at the geographic level. This approach adequately allows me to assess differences in MSAs that are driven jointly by race and other market characteristics, such as poverty levels, differences in population density, healthcare utilization norms, and healthcare facility and market characteristics. This helps to create a full picture of the place-based context under which disparities occur, such as the social determinants of health that may impact quality and access to care.

The MSA-level race measure is also large enough to ensure consistency between area of residence and healthcare use. Measuring racial composition with smaller geographic units may bias results if individuals have a propensity to seek healthcare outside of their neighborhood or county of residence. The U.S. has a longstanding history of housing policies dating back to Jim Crow laws and other racially-restrictive codes that led to institutionalized residential segregation. Although policy efforts such as the Fair Housing Act of 1968 aimed to eliminate housing discrimination, research suggests that 92.8% of the White population and 79.7% of the Black population continued to live in non-integrated neighborhoods as of 2015 (Massey, 2015; Reardon et al., 2008). This has spilled over into the provision of healthcare, where, in the case of childbirth, 75% of Black women give birth at hospitals that serve predominantly Black populations (Howell et al., 2016). A patient's hospital selection is a function of many factors; although proximity to one's home is important, research shows that individuals are willing to travel further for shorter wait times and better quality of care (Bühn, Holstiege, & Pieper, 2020; Finlayson, Birkmeyer, Tosteson, & Nease, 1999; Magee, Davis, & Coulter, 2003; Shalowitz, Nivasch, Burger, & Schapira, 2018; Varkevisser & Van Der Geest, 2007; Victoor, Delnoij, Friele, & Rademakers, 2012; Victoor et al., 2014). Further, because there is wide variation in the number of hospitals per capita, many residents may need to travel outside of their neighborhood or county for adequate access to care. Examining racial composition at the MSA-level ensures that there is concordance between the MSA in which a person lives and the hospital where they seek perinatal care.

The MSA-level race variable may also inform participation for future bundled payment programs. Assignment to prior programs, such as the Medicare Comprehensive Joint Replacement program, is designated at the MSA-level using areal-level healthcare costs, to determine which areas would benefit most from the incentive to reduce overall spending. Thus, examining variation in effects of APII by MSA-level racial composition is valuable in considering whether MSAs with a larger Black population would benefit from future enrollment with respect to disparities.

3.3.6. Control Group

The control group consists of twelve states that did not implement APII (California, Connecticut, Florida, Georgia, Indiana, Kentucky, Michigan, North Carolina, New Jersey, Texas, Virginia, and Wisconsin). States were selected on the basis of having at least one "Black" MSA and one "White" MSA in the MarketScan sample, to guarantee adequate representation of both race categories. This group also represents a diverse set of states across multiple regions, which may increase external validity.

I constructed the comparison group using a modified version of the multiple propensity score weighting approach proposed by Stuart et al. (2014), to ensure that control states are similar to treatment states across observable covariates (Stuart et al., 2014). This method sought to mitigate the concern that selection into treatment and control groups may be confounded by baseline characteristics, which can lead to biased estimates. This can occur if treatment and comparison groups vary in ways that impact trends over time, or within-group composition changes over time.

By matching treatment to control units on a set of shared factors, propensity scores aimed to replicate the pre-policy values of the outcome's determinants. The resulting weights helped to generate a synthetic sample in which treatment assignment is independent from measured baseline covariates (Austin & Stuart, 2015; Stuart et al., 2014, 2013; Zhang, Kim, Lonjon, & Zhu, 2019).

I fit two multinomial logistic regression models – one in the pre-period and the other in the post-period – to predict the probability of being assigned to the treatment group in each time frame. I adjusted for all covariates included in the main model, except for those that that could potentially be influenced by APII in the post-period. I also added interaction terms that multiply an indicator for Black MSA with each covariate. For observation i, the weight used in the analysis was obtained using the following calculation, where $E_1(X_i)$ is the probability that an observation is assigned to the treatment group, and $E_g(X_i)$ is the probability that the observation is assigned to their actual group:

$$w_i = \frac{E_1(X_i)}{E_g(X_i)} \tag{7}$$

By implementing propensity score weights without use of the outcome variables, I detach the study design from the analysis, and create the best opportunity for feasible, robust, and unbiased estimates (Stuart et al., 2014).

3.3.7. Empirical Strategy

The identification strategy is a difference-in-difference-in-differences (DDD) framework comparing Black and White MSAs in treatment and control states, pre-APII implementation (2010-2012) versus post-APII implementation (2013-2016). APII acts as a source of exogenous variation, as physicians in Arkansas are subject to mandatory bundled payments, while control states are not, leading to a quasi-experimental design. This approach is modeled on earlier work comparing average effects of APII on spending and utilization (C. Carroll et al., 2018).

In a standard difference-in-differences (DD) strategy, I would test for parallel pre-trends to indicate whether treatment states would have similar trends to the control group if the policy had not been implemented. The more complex DDD framework is protected in a sense, because the comparison of interest is the difference between Black versus White MSAs, rather than strictly focusing on differential changes in treatment versus control groups. Instead, following Paik et al. (2016), I compared pre-trends in disparities between Black versus White MSAs for each outcome in treatment and controls, to test whether the racial gap between treatment and control states would have evolved similarly over time in the absence of the policy (Paik, Black, & Hyman, 2016).

I estimated the impact of APII for Black and White MSAs using the following equation:

$$Y_{imst} = \beta_0 + \beta_1 \cdot \text{Treat}_s \cdot \text{Partial}_t \cdot \text{Black}_m + \beta_2 \cdot \text{Treat}_s \cdot \text{Partial}_t \cdot \text{White}_m + \beta_3 \cdot \text{Treat}_s \cdot \text{Full}_t \cdot \text{Black}_m + \beta_4 \cdot \text{Treat}_s \cdot \text{Full}_t \cdot \text{White}_m + \beta_5 \cdot \text{Treat}_s + \beta_6 \cdot \text{Full}_t + \beta_7 \cdot \text{Full}_t + \beta_8 \cdot \text{Treat}_s \cdot \text{Partial}_t + \beta_9 \cdot \text{Black}_m \cdot \text{Partial}_t + \beta_{10} \cdot \text{Treat}_s \cdot \text{Full}_t + \beta_{11} \cdot \text{Black}_m \cdot \text{Full}_t + \beta_{12} \cdot \text{Black}_m \cdot \text{Treat}_s + \zeta \cdot P_i + T_t + \vartheta \cdot Z_m + \mu \cdot V_s + \varepsilon_{imst} \quad [8]$$

In [8], Y_{imst} is the expected value of the outcome. It is indexed by the episode i, State-MSA *m*; in state *s*; at time *t*, which is representative of pre/post policy implementation. Treat is a binary variable that denotes the presence of the Arkansas Payment Improvement Initiative, while Partial and Full are binary variables that indicate the post-periods, including partial implementation in 2013 when APII was launched in Medicaid only, or full implementation from 2014-2016 when APII was active among both commercial and Medicaid insurers. Black_m and White_m, defined earlier, refer to MSAs with a high Black or White population, respectively. V_s and T_t are state and quarter-year fixed-effects, respectively. P_i is a vector of episode-level maternal characteristics. Z_m is a vector of time-varying State-MSA-level controls, which account for healthcare, demographic, and economic variables that may influence physician practice patterns. To account for covariance in standard errors between time periods by geographic areas, I clustered standard errors at the State-MSA level. All models were estimated using logistic regression.

The coefficients of interest were $\beta_1 - \beta_2$ and $\beta_3 - \beta_4$, which respectively, represent the aggregate effect of APII in Black versus White MSAs after partial and full implementation. I calculated these estimates using a linear combination. For high-value outcomes, a negative value for the difference in these coefficients indicates that White MSAs experienced greater quality gains than Black MSAs in Arkansas relative to controls. A positive value for the same estimates suggests that Black MSAs experiencing higher quality gains than White MSAs in Arkansas compared to controls. In contrast, for low-value outcomes, a positive value for the difference in the coefficients of interest indicates a greater improvement in quality for White MSAs compared to Black MSAs in Arkansas versus controls. Interpretation of each linear combination for specific outcomes can be found in Figure 3-3.

3.3.8. Robustness Checks

I pursued a variety of robustness checks. First, I repeated the analyses without multiple group propensity score weights. Second, I re-ran analyses using alternate comparison groups by dropping one control state at a time. Third, I re-estimated the models with MSA-level racial composition as a continuous measure, to examine whether results were robust to variable specification and the cutoff point for classifying Black and White MSAs. Finally, I re-ran analyses with episodes that were excluded from APII reimbursement to assess whether results varied based on patient risk level.

3.4. Results

3.4.1. Descriptive Statistics

Table 3-1 summarizes sample characteristics in Black and White MSAs among treatment and control states, before and after implementation of APII. Differences between treatment and control states are relatively small in magnitude during the pre-policy period across maternal characteristics, and MSA-level healthcare and economic factors, in both Black and White MSAs. Prior to APII, Black MSAs in both treatment and control groups had a higher proportion of HMO enrollees and births with a length of stay over 4 days. Black MSAs in both groups also had more Medicaid patients, more non-profit and teaching hospitals, more hospitals that provide obstetric care, more hospital beds per 1,000, a greater number of FQHCs, and a higher density of primary care physicians. On average, the population in Black MSAs was more educated and less likely to be uninsured or impoverished compared to White MSAs in both treatment and control groups; this population was also more likely to be unemployed. These patterns are consistent with a higher concentration of Black residents in urban areas (Leibbrand, Massey, Alexander, Genadek, & Tolnay, 2020). These areas also tend to include a higher concentration of universities and academic medical centers, suggesting that because these institutions are more likely to hire workers with higher education, the overall population is likely to be more educated and of higher socioeconomic status.

There is little evidence of differential changes in the sample after implementation of APII. The gap in percent of HMO enrollees, length of stay over 4 days, as well as Medicaid share decreased, while the difference in the percent of non-profit hospitals increased. Average percent cost sharing and education level rose across all groups, while the uninsurance and unemployment rates dropped. In general, gaps between groups declined. For example, the difference in the percent of mothers 35 or older was 5.76% and 5.23% in Black and White MSAs, respectively, in the pre-period, but these differences were reduced to 2.91% and 0.20% in the post-period. Otherwise, maternal, healthcare, demographic, and economic characteristic evolved similarly over time.

Table 3-2 and Table 3-3 summarize unadjusted rates of quality measures, pre- and post-APII. For high-value quality measures, Black MSAs in Arkansas had higher unadjusted levels of quality at baseline than White MSAs. For example, the Group B Strep screening rate in the pre-period for Black MSAs was 80.43% compared to 64.06% in White MSAs. Similarly, the rate of HIV screenings prior to APII among Black MSAs was 78.02%, and 59.62% among White MSAs. In the post-period, unadjusted quality rates improved in both Black and White MSAs. However, quality improvement was larger in White MSAs than in Black MSAs; in fact, White MSAs had higher unadjusted quality rates for five of the six prenatal screening measures in the post-period. In the first year after APII was implemented, the Group B Strep screening rate in White MSAs was

82.04% compared to 78.76% in Black MSAs. For HIV screenings, the increase in White MSAs was even larger in the same time frame, with an unadjusted rate of 74.85% compared to 71.50% in Black MSAs. In control states, the difference in unadjusted quality rates between Black and White MSAs was less significant. In the pre-policy period, five of the six prenatal screening measure rates were higher in White MSAs compared to Black MSAs, but differences were minor. For example, the A. Bact. screening rate was 89.32% in Black MSAs and 90.44% in White MSAs, representing relatively similar baseline quality rates across these measures. In the post-APII period, quality rates improved to a small extent in both Black and White MSAs, but to a lesser degree than Black and White MSAs in the treatment group (Table 3-2 and Table 3-3).

For low-value care, in which a lower rate is considered better, Black MSAs had poorer unadjusted quality rates at baseline compared to White MSAs, in both treatment and control groups. In the pre-APII period, the low-risk c-section rate in Black MSAs in Arkansas was 27.35% compared to 17.76% in White MSAs. In control states, Black MSAs experienced a low-risk c-section rate of 21.09%, relative to 16.18% in White MSAs. In the post-APII period, quality remained worse among Black MSAs. In 2013, the unadjusted low-risk c-section rate in Arkansas was 28.50% in Black MSAs compared to 13.17% in White MSAs. In the same year, control states showed greater improvements among Black MSAs, with a rate of 18.36% relative to 18.12% in White MSAs (Table 3-3).

Overall, these patterns suggest that Black MSAs experienced smaller improvements in both high- and low-value quality of care compared to White MSAs after partial implementation of APII. However, it is important to note that baseline quality rates were higher for prenatal and postpartum measures among Black MSAs, suggesting that White MSAs had greater room for improvement after the implementation of APII.

3.4.2. Validity of Study Design

To strengthen the assessment of covariate balance in Table 3-1, I evaluated the propensity score weights to ensure that there were no systematic differences in characteristics between treatment and controls in the pre- and post- periods. Table 3-5 shows the standardized mean differences (SMDs) between treatment and control groups before and after applying the propensity score weights. The threshold for an acceptable SMD is somewhat subjective, but in general, the consensus is that a covariate is considered balanced if the SMD is below 0.25, and ideally, below 0.10 (Austin & Stuart, 2015; Stuart et al., 2014, 2013; Zhang et al., 2019). Initially, the pre- and post-policy samples had an average SMD of 0.52 and 0.34, respectively, with only 47% of covariates yielding an SMD below 0.25. After constructing the propensity score weights, the mean SMD was reduced to 0.16 and 0.09 in the pre- and post-periods, respectively. In addition, 79% of covariates in the pre-period and 89% in the post-period had SMDs below 0.25, with the majority of SMDs below 0.10. Implementing propensity score weights significantly improved covariate balance in the sample. The final sample shows relatively small differences in characteristics between treatment and control units.

I also tested the DDD assumption that trends in racial disparities between treatment states and control states would remain similar in the absence of the treatment policy. Although this cannot be tested directly, I analyzed differences in pre-policy trends between Black and White MSAs in treatment and control groups, to assess validity of the analysis. Figure 3-1 plots unadjusted quarterly means of the outcomes in treatment and control states for Black and White MSAs separately, where the red lines represent trends in the treatment group and the blue lines represent outcome trends in the control group. Visual inspection suggests that all of the outcomes had similar gaps in pre-policy trends between Black and White MSAs in Arkansas and control states (Figure 3-1). To verify that the disparity in trends was statistically similar, I ran formal regressions similar to the main model, with inclusion limited to the pre-policy period (Table 3-4). The coefficient of interest was the interaction between treatment, Black MSA, and a linear quarter-year time trend. I found no statistically significant differences in trends, suggesting that the identification strategy is valid.

3.4.3. Effect of the Arkansas Payment Improvement Initiative on Supply of High-Value Care

In Table 3-2 and Table 3-3, I study the effect of APII on supply of high-value care in Black versus White MSAs. I find that under partial APII implementation, Black MSAs exhibited smaller improvements in prenatal screenings and postpartum visits compared to White MSAs under APII. There is a statistically significant decrease in the log odds of receiving four of the six prenatal screenings in Black MSAs relative to White MSAs after partial APII implementation, in Arkansas versus controls (Table 3-2). This includes receipt of screenings for HIV (log odds: -0.622; standard error (SE): 0.274; p-value<0.05), Hepatitis B (log odds: -0.587; SE: 0.297; p-value<0.05), A. Bact. (log odds: -0.573; SE: 0.324; p-value<0.10), and Gestational Diabetes, (log odds: -0.557; SE: 0.308; p-value<0.10). Compared to the pre-policy period, the log odds of receiving a timely postpartum visit after partial implementation was significantly lower among Black MSAs compared to White MSAs, in Arkansas versus control states (log odds: -0.604; SE: 0.273; p-value<0.05) (Table 3-3). Because prenatal screening and postpartum visit rates were increasing across all groups, this result can be interpreted as a smaller increase among Black MSAs versus White MSAs in the treatment group compared to controls.

After APII had been fully implemented in both commercial and Medicaid markets, only two of the six prenatal screenings showed disproportionate gains among White MSAs compared to Black MSAs relative to the pre-policy period. Gestational diabetes (log odds: -1.450, SE: 0.389; p-value<0.01) and Group B Strep (log odds: -0.773; SE: 0.285; p-value<0.01) showed a statistically significant decrease in Black MSAs relative to White MSAs. Four of the six prenatal screenings and receipt of timely postpartum care showed no differences in the effect of APII on quality of care by MSA-level racial composition.

These results suggest that both Black and White MSAs experienced quality gains in high-value care, but Black MSAs took more time to see improvements. After partial implementation, there were greater quality gains in prenatal and postpartum care in White MSAs relative to Black MSAs in the treatment group compared to controls. All prenatal screenings with significant results, except for HIV, were measures not linked to the payment incentive. This finding indicates that physicians in White MSAs were able to increase quality for high-value care measures, regardless of whether they were tied to payment, while physicians in Black MSAs had relatively more success improving quality for measures linked to reimbursement. This potentially suggests that the payment incentive was successful in encouraging physician behavior change in physician populations in both Black and White MSAs. Based on this result, it is possible that tying quality directly to reimbursement has potential to counteract disparities. After full implementation, only two measures showed persistently greater quality gains in White MSAs relative to Black MSAs, suggesting that Black MSAs were eventually able to close the quality gap.

3.4.4. Effect of the Arkansas Payment Improvement Initiative on Supply of Low-Value Care

I also assess whether there is an association between APII and an MSA's racial composition in low-value care (Table 3-3). Relative to the control group, I find that the log odds of receiving a low-risk c-section increased by 48.4% in Black MSAs compared to White MSAs in the treatment group after partial implementation in 2013 (log odds: 0.484; SE: 0.239; p-value<0.05). There is a smaller, insignificant decrease in the log odds of receiving a low-risk c-section in Black MSAs after full implementation (log odds: -0.202; SE: 0.196; p-value>0.10). These results are consistent with treatment patterns observed for high-value outcome measures.

3.4.5. Place-Based Disparities

In Figure 3-2, I explore the change in disparities between Black versus White MSAs in Arkansas compared to controls through an event-study analysis. For high-value care outcomes, I add to the main regression in [8] a three-way interaction between indicators for the treatment, White MSA, and year, for each year in the study period. To ensure consistency in interpretation, I use a different interaction term for low-value care outcomes; this inverts the direction of the coefficient since a lower rate is considered better for these measures. I multiply indicators for treatment, Black MSA, and year. Thus, for all outcome measures, a positive coefficient indicates a larger disparity between Black and White MSAs in Arkansas compared to controls, and a negative coefficient shows a smaller disparity. Figure 3-2 shows the difference in adjusted quality between Black and White MSAs in Arkansas versus control states, by year. The 2013 coefficient, depicted by yellow shading in the figure, represents the racial disparity after partial APII implementation, and the 2014 through 2016 coefficients, depicted by red shading in the figure, represent the racial disparity after full APII implementation. Taken together, this analysis shows a year-over-year trend in how place-based racial disparities evolve across the study period. This analysis is intended as a supplement to the main DDD analysis because it addresses how adjusted quality rates differ in the pre-period compared to the post-period. In contrast, results described in Table 3-2 and Table 3-3 highlight the change in the level of quality across each outcome in Black and White MSAs, in the treatment relative to control group, but do not take into account the differences in baseline quality rates.

I find some evidence of rising disparities. Results in Figure 3-2 show an increase in place-based disparities for low-risk c-sections in Arkansas compared to controls after partial implementation of APII. The difference in the low-risk c-section rate between Black and White MSAs in Arkansas was 72% higher than the equivalent gap in control states (p-value<0.01) in 2012, and 115% after partial APII implementation in 2013 (p-value<0.01). By the end of the post-policy period in 2016, the disparity returned to its initial magnitude.

Results were mixed for measures of high-value care. For prenatal screenings, place-based disparities in Arkansas during the pre-period were smaller than or equal to those in control states. Three measures (Group B. Strep, Gestational Diabetes, and A. Bact. screenings) showed a significantly higher disparity in Arkansas versus controls at the end of the study period (Group B. Strep – log odds: 1.34; p-value<0.05; Gestational Diabetes – log odds: 1.10; p-value<0.01; A. Bact. – log odds: 1.61; p-value<0.01). Two screening measures (HIV and Hepatitis B) experienced a temporary increase in disparities in 2013. Respectively, these measures showed a 59% (p-value<0.10) and 54% (p-vale<0.05) higher gap in place-based racial disparities in Arkansas relative to controls, but these disparities dissipated by 2014. Disparities were no different in Arkansas versus controls for Chlamydia screenings. For postpartum care, disparities remained lower in Arkansas versus controls over the entire study period.

Thus, I observe rising disparities after APII implementation across the majority of outcome measures. Disparities tended to be higher in Arkansas compared to control states in the short-term after partial implementation of APII. However, while both Black and White MSAs experienced quality improvement during the study period, racial disparities in Arkansas did not differ significantly from those in control states by the time APII had been implemented in both Medicaid and commercial markets.

3.4.6. Robustness Checks

Results are robust to alternate specifications and samples. This includes repeating analyses without multiple group propensity score weights (Table 3-6) and with MSA-level racial composition as a continuous, rather than binary, measure (Table 3-9). I also test alternate samples by re-running models with alternate comparison groups (Table 3-9), and with inclusion of episodes that are excluded under APII reimbursement (Table 3-7). Since results are consistent across each of these models, it suggests that inferences from the main models can be generalized more broadly.

3.5. Discussion

This paper studies the effect of a 2013 mandatory, multi-payer bundled payment program in Arkansas (APII) on place-based racial disparities in perinatal quality of care. Analyses compare how APII impacted the supply of low-value and high-value prenatal, intrapartum, and postpartum services in the commercial market in predominantly Black versus White metropolitan statistical areas (MSAs), in Arkansas relative to states with no change in reimbursement. I examine these changes in two stages of policy implementation: after the policy was implemented among Medicaid births only (partial implementation), and after it was enacted in Medicaid and commercial markets (full implementation). Assessing place-based, rather than individual racial disparities, enables me to study how the interaction between race and socioeconomic drivers contributes to variation in program effects across markets. The study uses an MSA-level measure of racial composition to create a full picture of the place-based context under which disparities occur, such as the social determinants of health that may impact quality and access to care, and to ensure concordance between area of residence and healthcare use.

Findings show that after partial implementation of APII, MSAs with a high proportion of Black residents were less likely to experience improvements in quality than MSAs with a high proportion of White residents. Specifically, for four of six prenatal screening measures (HIV, Hepatitis B, Gestational Diabetes, and A. Bact.), Black MSAs had a 55.7% to 62.2% lower probability of improving quality relative to White MSAs, in Arkansas versus control states. I found similar patterns for receipt of timely postpartum care and low-risk c-sections, with Black MSAs being 60.4% and 48.4% less likely to improve quality for these services, respectively. There was no strong evidence that disproportionate gains among White MSAs persisted after full implementation, with only two measures (Group B Strep and Gestational Diabetes screenings) displaying long-term improvements in White relative to Black MSAs. This study also compares the differential effects of episode-based payment between Black and White MSAs on quality measures that are linked to payment versus those that are not. Measures linked to gainsharing were less likely to show differential improvement across Black and White MSAs than those not tied to reimbursement. However, quality rates for all measures increased in both Black and White MSAs across the study period.

Results are consistent with prior work on the effects of value-based incentives on vulnerable populations. Existing studies offer conflicting findings. Some work shows that the introduction of non-FFS reimbursement leads to smaller gains in quality improvement for providers that treat a high proportion of marginalized patients (Chaiyachati et al., 2018; Colla et al., 2012;

Figueroa et al., 2018; Gaskin et al., 2018; McWilliams et al., 2017; Song et al., 2017), especially when the program incorporates a financial incentive with downside risk (e.g. penalizes physicians for poor quality or high spending) (Chaiyachati et al., 2018; Figueroa et al., 2018; Gaskin et al., 2018; Gilman et al., 2014; Gilman, Hockenberry, et al., 2015; Gu et al., 2014; Joynt & Jha, 2011, 2013; Joynt Maddox, 2018; Joynt Maddox, Reidhead, Qi, et al., 2019). Other studies find no association between financial incentives and changes in racial and socioeconomic disparities (Hsu et al., 2020; Stone, 2020). Results in this study provide some reconciliation for these disparate findings. I observe that while both Black and White areas experience quality improvement as a result of APII, areas with more Black residents experience smaller gains in quality in the first year after implementation. However, in subsequent years, these differences dissipate. It is possible that although quality is steadily improving in all areas, racial disparities temporarily widen at the program's onset, but are reduced to pre-policy levels over longer periods. These findings build on previous work by examining differential quality improvement among Black and White MSAs in response to a mandatory, non-Medicare bundled payment program.

If results can be generalized, then findings suggest that mandatory bundled payments increase quality within both Black and White MSAs, yet areas with a high proportion of Black residents experience less initial improvement. The most plausible explanation for these findings is that geographic areas with lower quality at baseline make the largest initial gains because they have greater room for quality improvement. Contrary to expectations, Black MSAs have better quality in the pre-policy period, for all outcome measures except for low-risk c-sections. This may be attributable to systematic differences in health systems; areas with a high proportion of Black residents tend to be large metropolitan areas, with more non-profit and academic medical centers. Theoretically, these institutions place greater emphasis on reputation and quality in the absence of a financial incentive (Horwitz, 2005; Lakdawalla & Philipson, 1998, 2006; Newhouse, 1970). It is possible that given higher baseline quality rates, Black MSAs experience "topping out" of quality measures upon APII implementation (Damberg et al., 2014; Golding, Nicola, Duszak, &

Rosenkrantz, 2020; McGlynn et al., 2003; Rathi, 2021). If this is the case, it suggests that the bundled payment program is functioning as designed; it incentivizes physicians with the lowest quality at baseline to change their behavior, leading to higher quality overall. Nonetheless, this study shows some evidence that White MSAs in Arkansas surpass Black MSAs on select quality measures, suggesting that Black MSAs may have more room for improvement, after all.

Another potential explanation is that physicians in White MSAs can more easily change their behavior in anticipation of the program's full implementation. Differences in physician behavior across Black and White MSAs is a function of marginal patient costs; providers in White MSAs may have the ability to make early modifications in quality of care, due to less complex patient populations, requiring smaller average investments in these patients. This is potentially conceivable given that measures with the most significant gains in White MSAs relative to Black MSAs are those not linked to gainsharing (Hepatitis B, Gestational Diabetes, and Asymptomatic Bacteriuria). Physicians in Black MSAs may face higher marginal patient costs to improve quality, so it makes sense that initially, they would prioritize the limited set of measures linked to gainsharing. One caveat is that Black MSAs tend to include a higher proportion of academic medical centers, which, on average, possess greater absolute healthcare resources due to higher payments from public and private payers. However, recent work has found an inverse relationship between hospital profits and academic medical center status, driven by disproportionately high spending. These facilities are more expensive than general acute care hospitals because they are tertiary care centers that attract severely ill, more expensive patients (Rosko et al., 2020).

This explanation potentially counters the earlier narrative that Black MSAs experience "topping out" of quality measures. It is important to note that urban areas with a greater share of academic medical centers may concurrently exert exceptional focus on quality in the absence of financial incentives, and experience cost-related challenges in improving quality for complex patient populations. Future work should continue to consider how areal-level racial and socioeconomic composition interact to create heterogeneity in episode-based payment program effects, and potential effects across different hospital types.

Alternately, these findings may be driven by the program's staggered rollout. APII was predominantly implemented in the Medicaid market in 2013, followed by the commercial market in 2014. Therefore, the initial quality gains in White MSAs relative to Black MSAs in the treatment group may represent a spillover from Medicaid to commercial enrollees. In the case of a spillover, it is expected that larger effect would be observed in the commercially insured population as the share of Medicaid patients rises. In this study, the opposite occurs, as Black MSAs in Arkansas have a higher share of Medicaid patients (21.25%) relative to White MSAs (12.84%), diminishing the likelihood of this explanation. Future analyses may benefit from exploring the mechanism for the main result.

As several states and payers continue to debate episode-based payments and other bundled payment policies, this analysis suggests that these incentives may be successful on average, but policymakers should consider the variation in effects. Specifically, APII led to differential short-term effects between Black and White MSAs; while overall quality seemed to rise, effects were significantly larger among MSAs with a smaller Black population, especially after partial implementation. This general pattern is consistent with a prior evaluation of the program's average effects, which also found larger (albeit, null) quality gains after partial implementation (C. Carroll et al., 2018). This difference may be due to masked heterogeneity leading to attenuation, or use of a different control group (Markovitz & Ryan, 2017). Notably, this study is the first to explore effects of multi-payer mandatory bundled payments on place-based racial disparities. Results contribute to a growing empirical literature on racial disparities and physician behavior, showing that long-term financial incentives may improve perinatal quality overall, but I find no evidence that this program narrows racial disparities.

Results have implications for efforts to expand bundled payment reforms. This analysis indicates that while bundled payments may be successful on some dimensions, there is no evidence that they mitigate racial disparities. This suggests a need for payment policies to more directly target racial equity. Prenatal screenings linked to gainsharing were less likely to show disproportionate gains among White MSAs, suggesting that tethering specific measures to payment may be an effective strategy for bolstering quality without intensifying disparities. There were smaller changes in disparities among prenatal screening measures tied to reimbursement, suggesting that directly linking a financial incentive to racial equity may be a promising approach to narrow disparities.

This result also offers evidence against risk-adjusting quality measures on the basis of race, which is a controversial topic. Proponents argue that risk-adjustment "levels the playing field" and encourages physicians to take on patients with different clinical and social profiles, as it avoids imposing penalties for high-risk patients (Cher, Ryan, Hoffman, & Sheetz, 2020; Glance et al., 2016; Hoffman, Hsuan, Braun, & Ponce, 2019; Hu, Gonsahn, & Nerenz, 2014; Joynt Maddox, Reidhead, Hu, et al., 2019; J. R. Martin et al., 2017; Roberts et al., 2018; Sills et al., 2017). Meanwhile, critics are concerned that risk-adjustment can mask differential treatment of patients according to racial and social characteristics, and may lower quality standards for disadvantaged patients (Anderson, Li, Romano, Parker, & Chang, 2016; Bynum & Lewis, 2018; Meddings et al., 2017; Tran, 2020). Results from this study suggest that outlining clear goals for physicians on dimensions racial equity (e.g. reducing disparities), rather than risk-adjusting for these differences, may be a central component of future bundled payment design. Future work exploring payment-related solutions to racial disparities, and potential implications of risk-adjustment, is critically important given current inequalities in healthcare.

This work also suggests that interventions focused on patient education and engagement may be helpful in accompanying the financial incentive. Gestational diabetes reflected the most persistent increase in place-based disparities. Notably, this is the most time intensive prenatal screening, as it involves fasting for eight hours before the test, followed by an initial blood glucose test, drinking a syrupy glucose solution, and then re-testing blood glucose levels one hour later.

Borderline or concerning results require a more intensive three-hour follow-up test. From a physician behavior perspective, widening disparities may be a function of unequal patient education and expectation-setting for patients in Black and White MSAs (e.g. racial discrimination). Patients in Black MSAs may experience disproportionately high barriers to accessing this type of care, including time constraints and health literacy issues. Prior work generally supports the notion that supply and demand side factors contribute to disparities in childbirth, but this study is the first to make strides towards isolating supply side drivers under an econometric framework.

This analysis has several limitations and suggests potential avenues for future work. First, I can only identify MSA-level, rather than patient-level, race data. This limits the ability to make inferences about within-MSA disparities. I argue that since Black and White residential areas remain largely segregated, and this pattern has spilled over to the healthcare sector, the MSA-level race statistic provides valuable information about whether differential trends in quality by MSA are related to patient race and neighborhood context. In the sample, Black MSAs are 25% Black, while White MSAs are only 4% Black, suggesting a wide enough gap in racial composition for differences in utilization to potentially stem from racial composition. Still, there are several possible explanations for the observed differential practice patterns by MSA, including unobserved differences between areas (e.g. cultural norms and preferences surrounding prenatal testing, physician characteristics, variation patient-physician communication, health literacy and access to information about appropriate treatment options, and demographic or family dynamics that impact behavior outside of the healthcare system).

Further, the MarketScan database does not identify individual payers or physicians. This blurs the distinction between physician and broader health system behavior changes and limits the ability to attribute the observed effects to physician behavior. It will be useful for future work to address this gap and examine these effects among individual physicians. Next, I focus on Arkansas payment policy in perinatal care, so results may not be generalizable to other clinical or geographic areas. From a demographic perspective, Arkansas may be poorly suited to exploring racial disparities, since it is small with a predominantly White population. Additional work is needed to strengthen this claim, potentially in more diverse states. Finally, the DDD design rests on the assumption that no unobserved factors contribute to the results. I address this concern by using a doubly robust propensity score weighting approach, supplemented with several robustness checks. However, this approach does not account for time-varying unobserved factors, which may weaken interpretation of results.

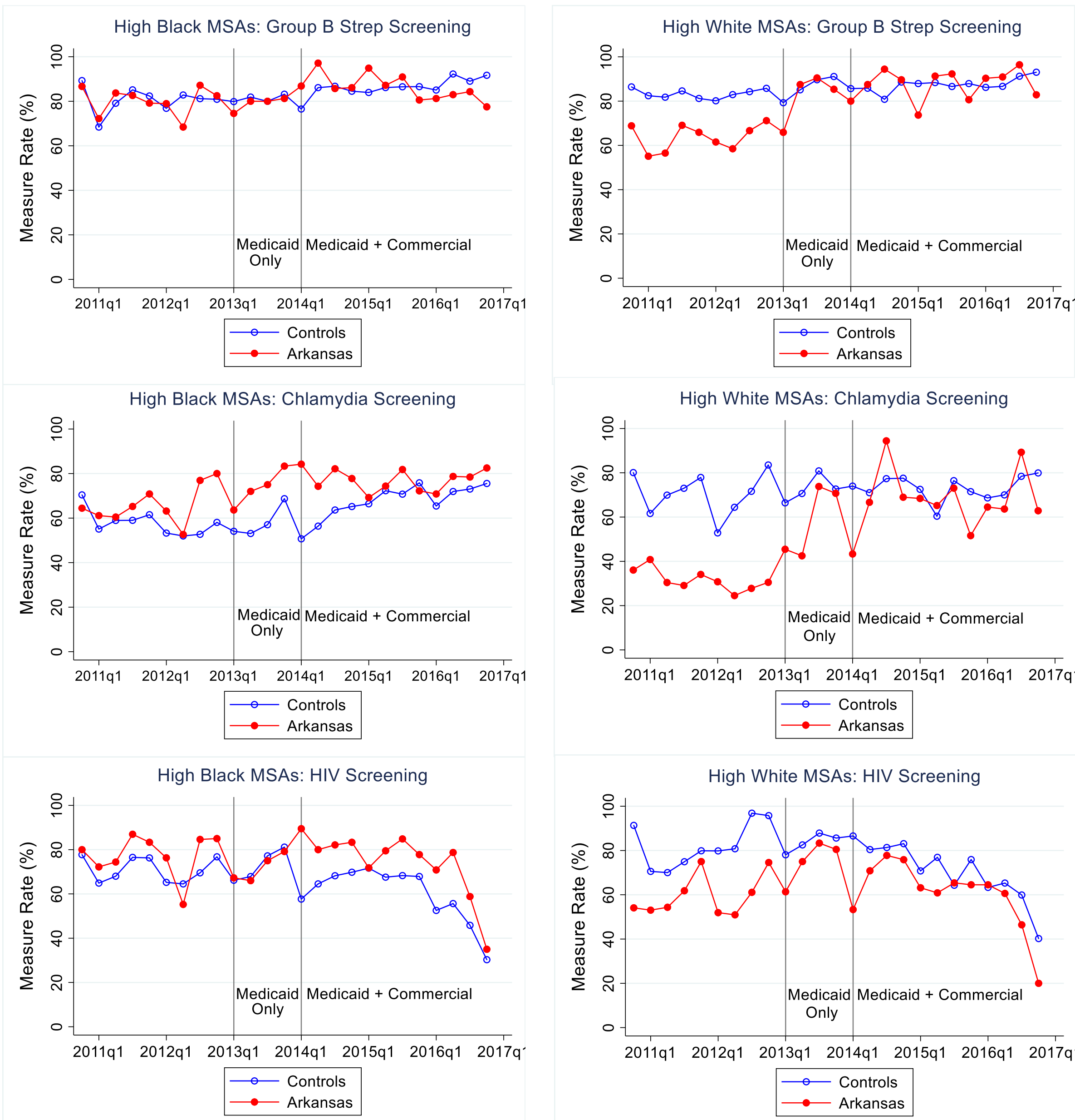
Non-FFS payment reforms are becoming increasingly salient, but there are remaining questions about how to design incentives that promote racial equity. I present evidence that current incentives lead to disproportionate gains in areas with more White residents in the short-term, but these gaps remain unchanged in the long-term. While results are specific to perinatal care, this study provides general guidance about whether bundled payments can serve as a tool for reducing racial disparities. Continuing to build an understanding of these incentives and the variation in their effects across populations, is imperative for future work.

3.6. Figures

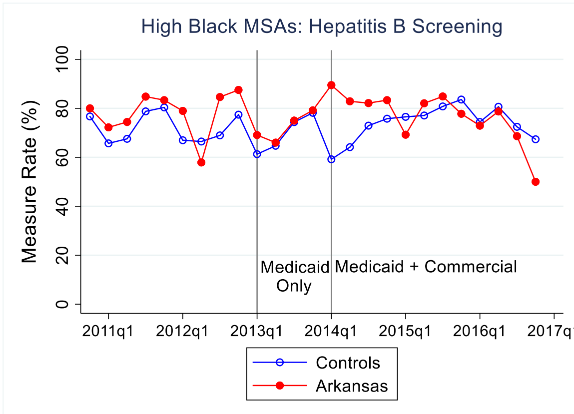
Figure 3-1 Trends in Prenatal, Intrapartum, and Postpartum Quality for Black versus White Metropolitan Statistical Areas, 2010-2016

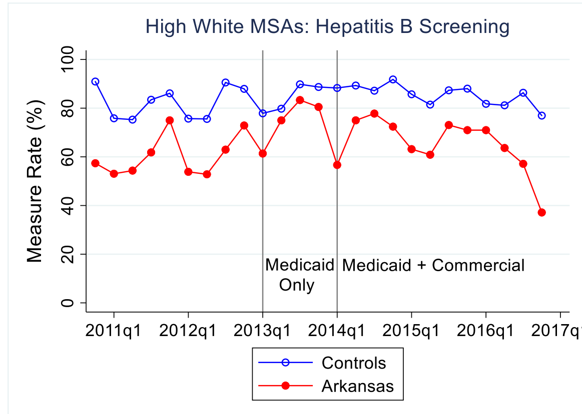
<u>Prenatal Screenings</u>

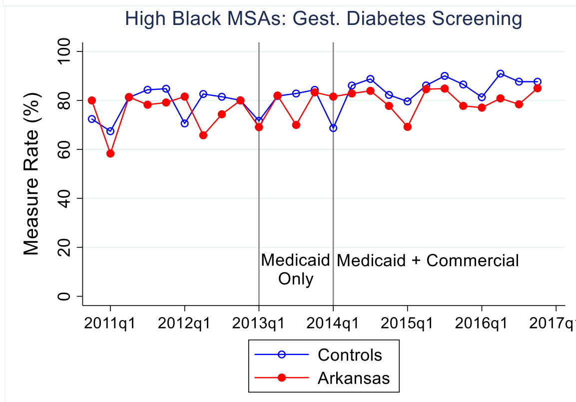
Linked to Gainsharing:

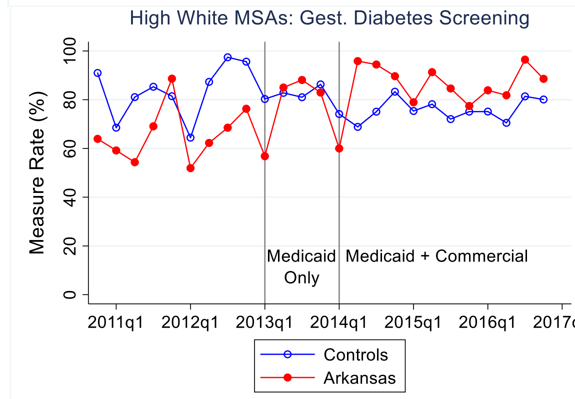


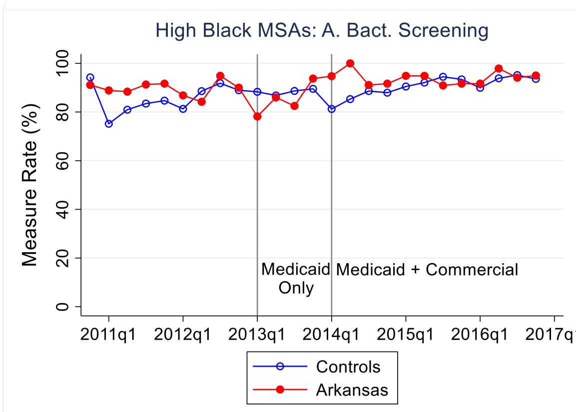
Not Linked to Gainsharing:

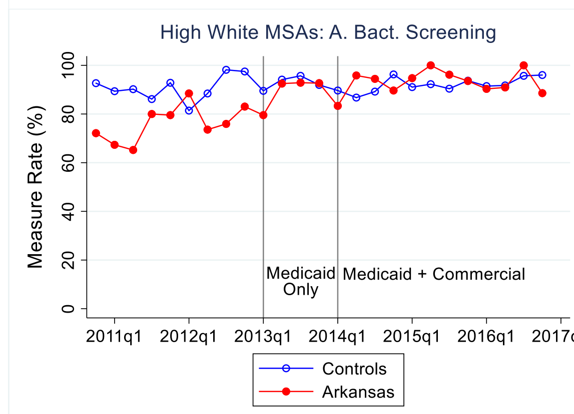




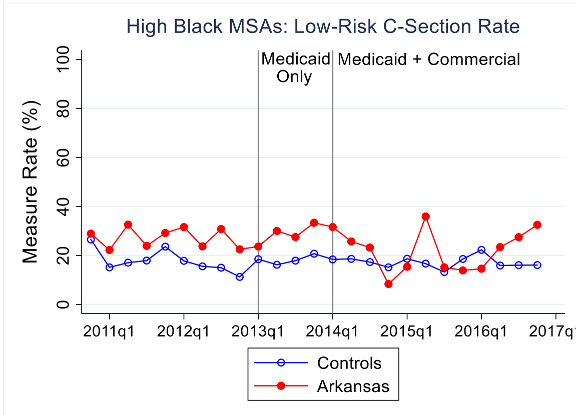


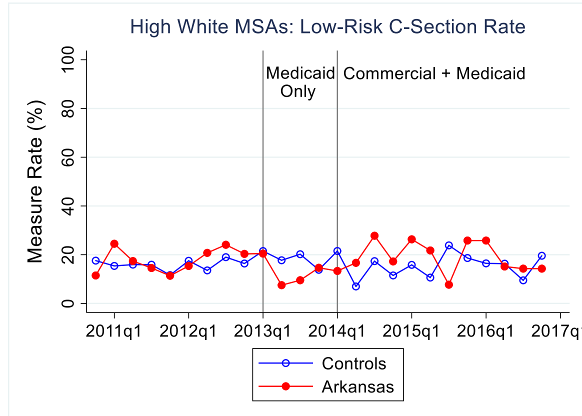




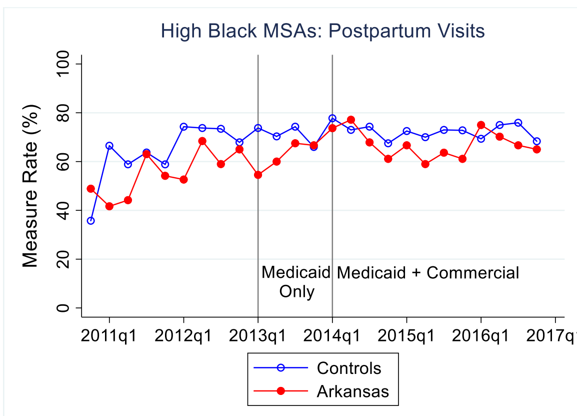


Intrapartum Care:





Postpartum Care:



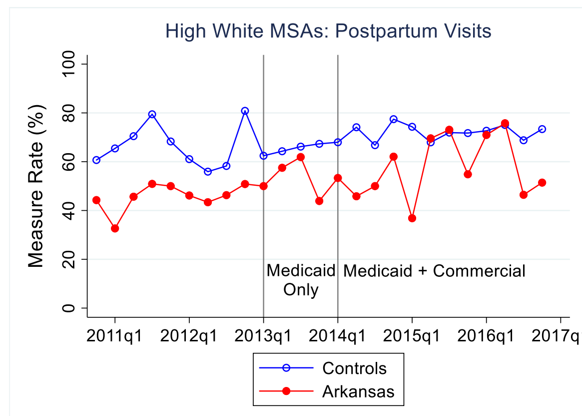
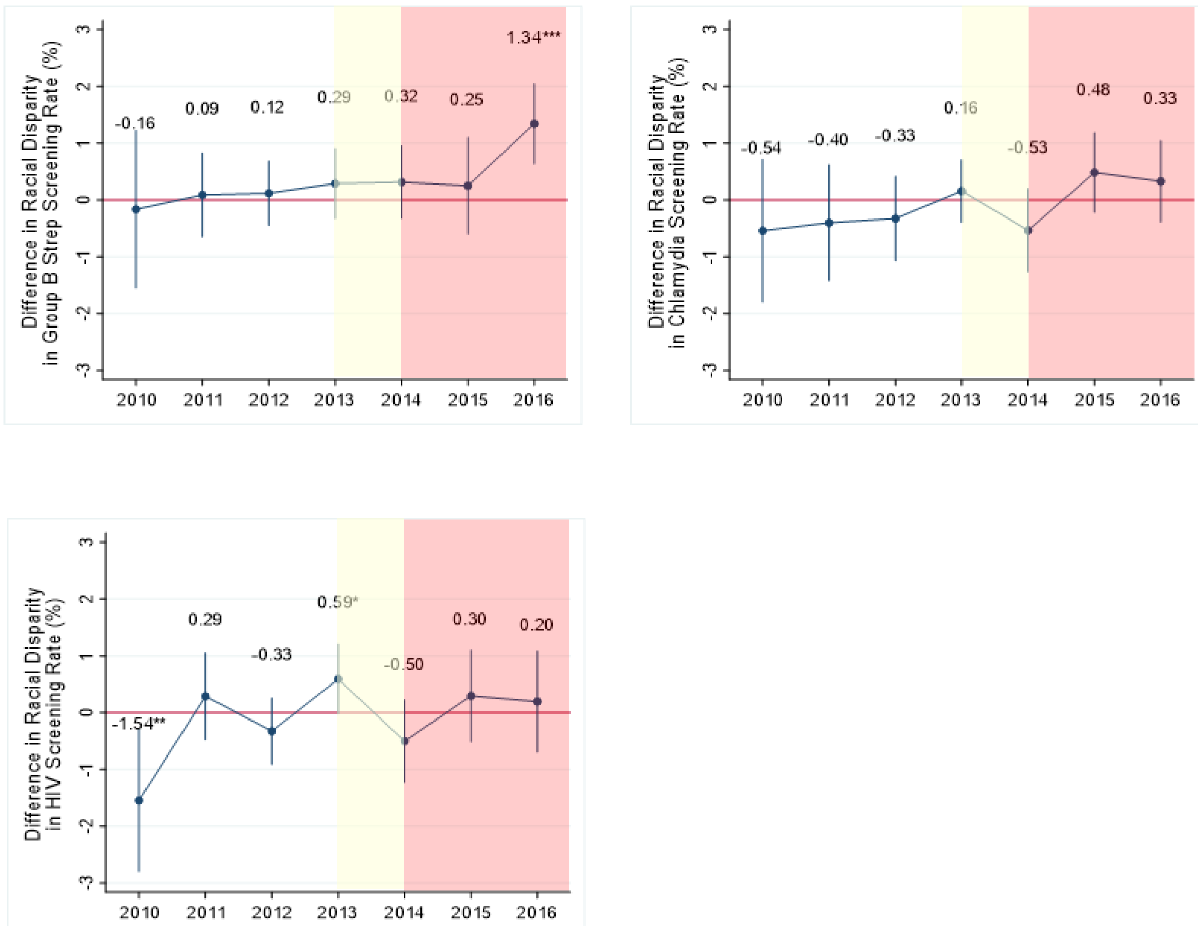


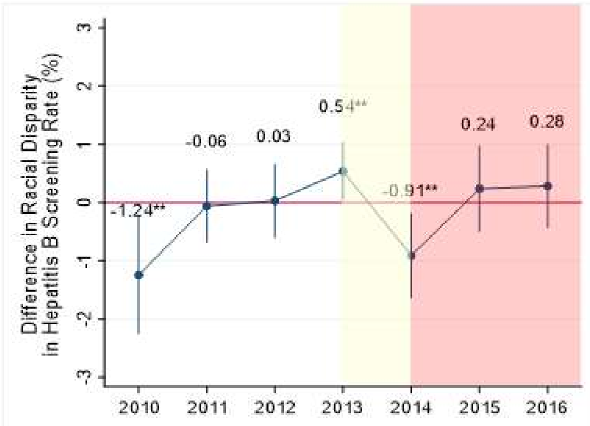
Figure 3-2 Change in Place-Based Racial Disparities Pre- versus Post-Implementation of the Arkansas Payment Improvement Initiative

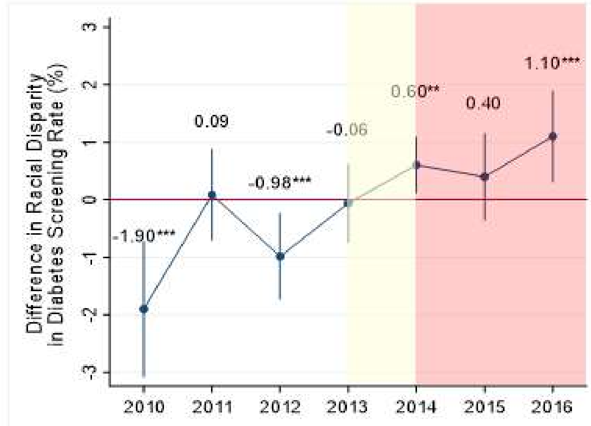
<u>Prenatal Screenings</u>

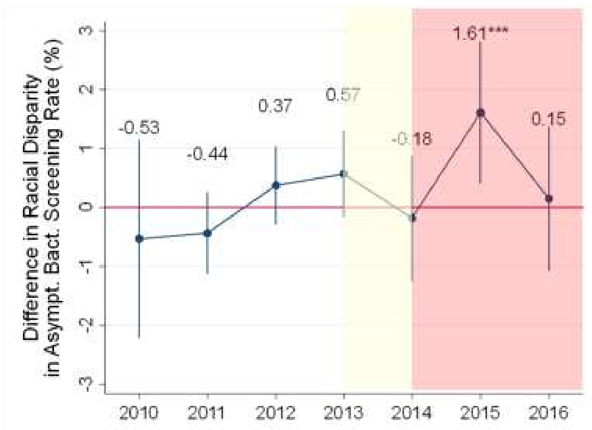
Linked to Gainsharing:



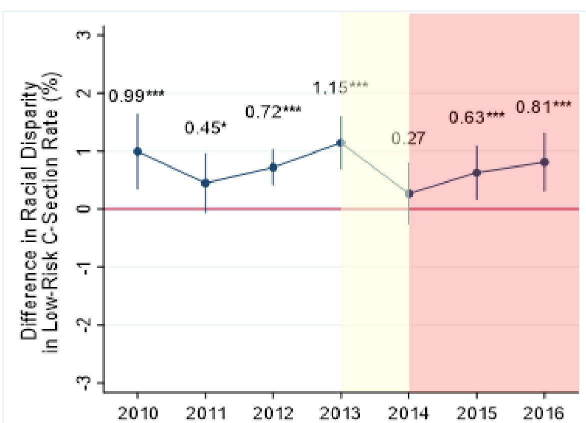
Not Linked to Gainsharing:



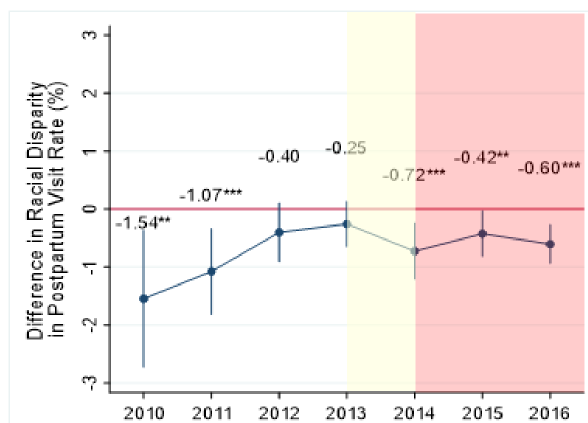




<u>Perinatal Care:</u>



<u>Postpartum Care:</u>



Notes: Sample estimates are from the Truven MarketScan claims, using data from births between 2010-2016. Coefficients are modeled using logistic regression and are displayed as log odds. Coefficients and 95% Confidence Intervals for the DDD results are shown, and can be interpreted as the perfect difference in Black versus White MSAs in Arkansas versus Controls in a given year. Standard errors are clustered at the State-MSA level. High Black population is determined by share of Black residents from 2010-2012, prior to policy implementation being above a threshold of 12.59%, which was the corresponding national average. Covariates include all variables in Table 3-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

3.7. Tables

Table 3-1 Summary Statistics of Arkansas and Control States, Pre- and Post- Implementation of the Arkansas Payment Improvement Initiative

	Pre APII (2010-2012)				**Post APII (2013-2016)**			
	Arkansas		Control States		Arkansas		Control States	
	Black MSAs	White MSAs	Black MSAs	White MSAs	Black MSAs	White MSAs	Black MSAs	White MSAs
Maternal Characteristics								
% Maternal Age 35+	9.92	14.38	15.68	9.15	11.00	15.38	13.91	15.18
% LOS > 4	1.88	0.63	1.33	0.51	1.88	1.01	2.02	0.98
Health Plan								
HMO	3.22	0.63	3.51	1.27	1.88	0.61	1.86	0.76
PPO	74.26	68.65	71.71	75.75	69.32	65.18	68.05	69.09
POS	13.94	4.65	11.36	11.19	11.87	5.87	8.65	9.16
High Deductible	8.58	16.08	13.42	7.94	16.93	28.34	21.44	21.00
% Cost Sharing	18.21	17.57	16.06	19.26	21.40	24.08	19.85	21.52
MSA-Level Healthcare Characteristics								
% Medicaid	21.25	12.84	17.44	13.73	16.96	13.17	14.60	13.24
% Hospitals Non-Profit	52.23	23.05	39.66	36.96	52.63	25.55	49.77	27.33
% Hospitals Provide OB Services	31.89	31.40	34.78	32.43	33.82	34.52	32.99	31.85
% Teaching Hospitals	11.23	0.00	11.87	4.74	11.28	0.00	9.20	4.27
# of FQHCs	34.39	14.13	48.77	86.57	39.53	41.03	141.72	267.07
Beds per 1,000	5.46	3.05	4.06	1.69	5.57	3.10	2.64	1.85
PCPs per 1,000	0.74	0.68	0.91	0.67	0.78	0.72	0.82	0.65
MSA-Level Demographic Characteristics								
% < HS Education	11.41	15.26	14.52	19.96	10.46	14.79	11.06	17.81
% > College Education	27.82	27.20	29.27	19.79	28.39	28.34	34.40	23.84
% Population Black								
MSA-Level Economic Characteristics								
% Uninsured	17.51	21.72	17.34	23.23	11.70	15.89	13.36	20.06
% Unemployed	7.18	6.20	8.87	7.42	5.52	4.71	5.89	5.88
% Poverty	15.17	18.26	15.61	17.29	15.30	16.78	14.45	16.98

Episodes	373	473	24,834	27,950	691	494	56,415	47,628
# of MSAs	3	3	45	93	3	3	46	95

Notes: Sample estimates are from the Truven MarketScan commercial claims database, using data from births during the study period. Healthcare characteristics are from the American Hospital Association Annual Survey, the Area Health Resource File, and the National Practitioner Data Bank. Demographic and economic characteristics are obtained from the U.S. Census American Community Survey.

Table 3-2 Prenatal Care: Variation in Effects of the Arkansas Payment Improvement Initiative Between Metropolitan Statistical Areas with a High Black Population versus a High White Population

	Linked to Gainsharing						**Not Linked to Gainsharing**					
	Group B Strep		Chlamydia		HIV		Hepatitis B		Gestational Diabetes		Asymptomatic Bacteriuria	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post												
Black	-0.164	-0.429	0.257	0.182	-0.344	-0.102	-0.413**	-0.385	0.140	-0.728**	-0.504**	-0.314
	(0.232)	(0.338)	(0.214)	(0.228)	(0.221)	(0.291)	(0.163)	(0.234)	(0.208)	(0.342)	(0.234)	(0.304)
White	0.177	0.344	0.685**	0.560*	0.278	0.025	0.174	-0.454	0.697***	0.721**	0.069	0.156
	(0.219)	(0.243)	(0.306)	(0.322)	(0.206)	(0.265)	(0.266)	(0.239)	(0.202)	(0.308)	(0.383)	(0.235)
Difference	-0.340	-0.773***	-0.428	-0.378	-0.622**	-0.127	-0.587**	0.069	-0.557*	-1.450***	-0.573*	-0.470
	(0.307)	(0.285)	(0.339)	(0.338)	(0.274)	(0.311)	(0.297)	(0.262)	(0.308)	(0.389)	(0.324)	(0.291)
N	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858
Dependent Variable Mean: %												
Black MSAs												
Pre-APII												
Treatment	80.43	80.43	66.22	66.22	78.02	78.02	78.55	78.55	75.87	75.87	89.81	89.81
Control	84.24	84.24	62.61	62.61	73.52	73.52	73.67	73.67	76.01	76.01	89.32	89.32
Post-APII												
Treatment	78.76	85.94	73.06	77.31	71.50	73.90	72.02	76.51	76.17	80.32	84.97	93.98
Control	81.20	86.26	58.33	68.10	73.24	61.22	69.76	74.79	80.13	84.97	88.36	90.97

White MSAs												
Pre-APII												
Treatment	64.06	64.06	31.50	31.50	59.62	59.62	60.47	60.47	65.96	65.96	76.32	76.32
Control	82.97	82.97	67.76	67.76	84.46	84.46	81.73	81.73	83.90	83.90	90.44	90.44
Post-APII												
Treatment	82.04	87.46	58.08	66.36	74.85	58.72	74.85	63.91	77.84	84.71	89.22	92.66
Control	86.72	86.11	73.03	74.05	83.84	76.58	84.52	87.12	83.74	75.17	92.93	91.01

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Black population is determined by share of Black residents from 2010-2012, prior to policy implementation being above a threshold of 12.59%, which was the corresponding national average. Covariates include all variables in Table 3-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3-3 Intrapartum and Postpartum Care: Variation in Effects of the Arkansas Payment Improvement Initiative Between Metropolitan Statistical Areas with a High Black Population versus a High White Population

	Intrapartum Care		**Postpartum Care**	
	Low-Risk C-Sections		Postpartum Visit Within 8 Weeks	
	2013	2014-2016	2013	2014-2016
Treat * Post				
Black	-0.084	-0.273**	-0.122	0.339*
	(0.150)	(0.131)	(0.168)	(0.189)
White	-0.568***	-0.072	0.482**	0.564***
	(0.176)	(0.179)	(0.234)	(0.217)
Difference	0.484**	-0.202	-0.604**	-0.225
	(0.239)	(0.196)	(0.273)	(0.258)
N	158,858	158,858	158,858	158,858
Dependent Variable Mean: %				
Black MSAs				
Pre-APII				
Treatment	27.35	27.35	55.23	55.23
Control	21.09	21.09	53.06	53.06
Post-APII				
Treatment	28.50	22.49	61.66	67.47
Control	18.36	17.10	71.14	72.49
White MSAs				
Pre-APII				
Treatment	17.76	17.76	45.67	45.67
Control	16.18	16.18	64.24	64.24
Post-APII				
Treatment	13.17	18.35	53.29	58.41
Control	18.12	15.18	65.23	71.43

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Black population is determined by share of Black residents from 2010-2012, prior to policy implementation being above a threshold of 12.59%, which was the corresponding national average. Covariates include all variables in Table 3-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

3.8. References

ACHI. (2015). *Arkansas Health Care Payment Improvement Initiative: 1st Annual Statewide Tracking Report*.

Agency for Healthcare Quality and Research. (2016). *AHRQ - Quality Indicators Technical Specifications*. Retrieved from https://www.qualityindicators.ahrq.gov/Modules/IQI_TechSpec_ICD10_v60.aspx

Allison, A. (2013). *Arkansas Medicaid Program Overview: SY2013*. Retrieved from https://humanservices.arkansas.gov/dms/DMS Public/Medicaid_Program_Overview_SFY2013.pdf

American Hospital Association. (2018). AHA Data Products: Annual Survey. Retrieved August 23, 2019, from https://www.aha.org/data-insights/aha-data-products

Anderson, J. E., Li, Z., Romano, P. S., Parker, J., & Chang, D. C. (2016). Should Risk Adjustment for Surgical Outcomes Reporting Include Sociodemographic Status? A Study of Coronary Artery Bypass Grafting in California. *Journal of the American College of Surgeons*, *223*(2), 221–230. https://doi.org/10.1016/j.jamcollsurg.2016.05.008

Arkansas Blue Cross Blue Shield. (2014). *Provider Manual: Payment Improvement Initiative*. Retrieved from www.ArkansasBlueCross.com.

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, *34*(28), 3661–3679. https://doi.org/10.1002/sim.6607

Bühn, S., Holstiege, J., & Pieper, D. (2020). Are patients willing to accept longer travel times to decrease their risk associated with surgical procedures? A systematic review. *BMC Public Health*, *20*(1), 253. https://doi.org/10.1186/s12889-020-8333-5

Bynum, J., & Lewis, V. (2018, November 1). Value-Based Payments and Inaccurate Risk Adjustment - Who Is Harmed? *JAMA Internal Medicine*, Vol. 178, pp. 1507–1508. https://doi.org/10.1001/jamainternmed.2018.4142

Carroll, C., Chernew, M., Fendrick, A. M., Thompson, J., & Rose, S. (2018). Effects of episode-based payment on health care spending and utilization: Evidence from perinatal care in Arkansas. *Journal of Health Economics*, *61*, 47–62. https://doi.org/10.1016/j.jhealeco.2018.06.010

Chaiyachati, K. H., Qi, M., & Werner, R. M. (2018). Changes to Racial Disparities in Readmission Rates After Medicare's Hospital Readmissions Reduction Program Within Safety-Net and Non-Safety-Net Hospitals. *JAMA Network Open*, *1*(7), e184154. https://doi.org/10.1001/jamanetworkopen.2018.4154

Cher, B. A. Y., Ryan, A. M., Hoffman, G. J., & Sheetz, K. H. (2020). Association of Medicaid Eligibility With Surgical Readmission Among Medicare Beneficiaries. *JAMA Network Open*, *3*(6), e207426. https://doi.org/10.1001/jamanetworkopen.2020.7426

Colla, C. H., Wennberg, D. E., Meara, E., Skinner, J. S., Gottlieb, D., Lewis, V. A., … Fisher, E. S. (2012). Spending Differences Associated With the Medicare Physician Group Practice Demonstration. *JAMA*, *308*(10), 1015. https://doi.org/10.1001/2012.jama.10812

Dafny, L. S. (2005, December). How do hospitals respond to price changes? *American Economic Review*, Vol. 95, pp. 1525–1547. https://doi.org/10.1257/000282805775014236

Damberg, C. L., Elliott, M. N., & Ewing, B. A. (2015). Pay-For-Performance Schemes That Use Patient And Provider Categories Would Reduce Payment Disparities. *Health Affairs*, *34*(1), 134–142. https://doi.org/10.1377/hlthaff.2014.0386

Damberg, C. L., Sorbero, M. E., Lovejoy, S. L., Martsolf, G. R., Raaen, L., & Mandel, D. (2014). Measuring Success in Health Care Value-Based Purchasing Programs: Findings from an Environmental Scan, Literature Review, and Expert Panel Discussions. *Rand Health Quarterly*, *4*(3), 9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28083347

Decker, S. L. (2013). Two-Thirds Of Primary Care Physicians Accepted New Medicaid Patients In 2011–12: A Baseline To Measure Future Acceptance Rates. *Health Affairs*, *32*(7), 1183–1187. https://doi.org/10.1377/hlthaff.2013.0361

Figueroa, J. F., Zheng, J., John Orav, E., Epstein, A. M., & Jha, A. K. (2018). Medicare program associated with narrowing hospital readmission disparities between black and white patients. *Health Affairs*, *37*(4), 654–661. https://doi.org/10.1377/hlthaff.2017.1034

Finlayson, S. R. G., Birkmeyer, J. D., Tosteson, A. N. A., & Nease, R. F. (1999). Patient preferences for location of care: Implications for regionalization. *Medical Care*, *37*(2), 204–209. https://doi.org/10.1097/00005650-199902000-00010

Gaskin, D. J., Zare, H., Vazin, R., Love, D., & Steinwachs, D. (2018). Racial and Ethnic Composition of Hospitals' Service Areas and the Likelihood of Being Penalized for Excess Readmissions by the Medicare Program. *Medical Care*, *56*(11), 934–943. https://doi.org/10.1097/MLR.0000000000000988

Getahun, D., Strickland, D., Lawrence, J. M., Fassett, M. J., Koebnick, C., & Jacobsen, S. J. (2009). Racial and ethnic disparities in the trends in primary cesarean delivery based on indications. *American Journal of Obstetrics and Gynecology*, *201*(4), 422.e1-422.e7. https://doi.org/10.1016/j.ajog.2009.07.062

Gilman, M., Adams, E. K., Hockenberry, J. M., Wilson, I. B., Milstein, A. S., & Becker, E. R. (2014). California safety-net hospitals likely to be penalized by ACA value, readmission, and meaningful-use programs. *Health Affairs*, *33*(8), 1314–1322. https://doi.org/10.1377/hlthaff.2014.0138

Gilman, M., Hockenberry, J. M., Adams, E. K., Milstein, A. S., Wilson, I. B., & Becker, E. R. (2015). The financial effect of value-based purchasing and the hospital readmissions reduction program on safety-net hospitals in 2014: A cohort study. *Annals of Internal Medicine*, *163*(6), 427–436. https://doi.org/10.7326/M14-2813

Glance, L. G., Kellermann, A. L., Osler, T. M., Li, Y., Li, W., & Dick, A. W. (2016). Impact of risk adjustment for socioeconomic status on risk-adjusted surgical readmission rates. *Annals of Surgery*, *263*(4), 698–704. https://doi.org/10.1097/SLA.0000000000001363

Golding, L. P., Nicola, G. N., Duszak, R., & Rosenkrantz, A. B. (2020). The Quality Measure Crunch: How CMS Topped Out Scoring and Removal Policies Disproportionately Disadvantage Radiologists. *Journal of the American College of Radiology*, *17*(1), 110–117. https://doi.org/10.1016/j.jacr.2019.08.014

Gu, Q., Koenig, L., Faerberg, J., Steinberg, C. R., Vaz, C., & Wheatley, M. P. (2014). The medicare hospital readmissions reduction program: Potential unintended consequences for hospitals serving vulnerable populations. *Health Services Research*, *49*(3), 818–837. https://doi.org/10.1111/1475-6773.12150

Hoffman, G. J., Hsuan, C., Braun, T., & Ponce, N. (2019). Health Equity and Hospital Readmissions: Does Inclusion of Patient Functional and Social Complexity Improve Predictiveness? *Journal of General Internal Medicine*, *34*(1), 26–28. https://doi.org/10.1007/s11606-018-4635-z

Horwitz, J. R. (2005). Making profits and providing care: Comparing nonprofit, for-profit, and government hospitals. *Health Affairs*, *24*(3), 790–801. https://doi.org/10.1377/hlthaff.24.3.790

Howell, E. A., Egorova, N. N., Balbierz, A., Zeitlin, J., & Hebert, P. L. (2016). *Site of delivery contribution to black-white severe maternal morbidity disparity*. https://doi.org/10.1016/j.ajog.2016.05.007

HRSA. (2019). Area Health Resources Files. Retrieved September 11, 2019, from https://data.hrsa.gov/topics/health-workforce/ahrf

Hsiang, W. R., Lukasiewicz, A., Gentry, M., Kim, C. Y., Leslie, M. P., Pelker, R., … Wiznia, D. H. (2019). Medicaid Patients Have Greater Difficulty Scheduling Health Care Appointments Compared With Private Insurance Patients: A Meta-Analysis. *Inquiry : A Journal of Medical Care Organization, Provision and Financing*, *56*. https://doi.org/10.1177/0046958019838118

Hsu, H. E., Wang, R., Broadwell, C., Horan, K., Jin, R., Rhee, C., & Lee, G. M. (2020). Association Between Federal Value-Based Incentive Programs and Health Care-Associated Infection Rates in Safety-Net and Non-Safety-Net Hospitals. *JAMA Network Open*, *3*(7), e209700. https://doi.org/10.1001/jamanetworkopen.2020.9700

Hu, J., Gonsahn, M. D., & Nerenz, D. R. (2014). Socioeconomic status and readmissions: Evidence from an urban teaching hospital. *Health Affairs*, *34*(5), 778–785. https://doi.org/10.1377/hlthaff.2013.0816

Johns Hopkins. (2016). Institute for Clinical and Translational Research: Truven Health MarketScan Research Databases. Retrieved September 11, 2019, from https://ictr.johnshopkins.edu/programs_resources/programs-resources/research-participant-recruitment-and-retention/truven-health-marketscan-research-databases/

Joynt, K. E., & Jha, A. K. (2011). Who has higher readmission rates for heart failure, and why? Implications for efforts to improve care using financial incentives. *Circulation: Cardiovascular Quality and Outcomes*, *4*(1), 53–59. https://doi.org/10.1161/CIRCOUTCOMES.110.950964

Joynt, K. E., & Jha, A. K. (2013, January 23). Characteristics of hospitals receiving penalties under the hospital readmissions reduction program. *JAMA - Journal of the American Medical Association*, Vol. 309, pp. 342–343. https://doi.org/10.1001/jama.2012.94856

Joynt Maddox, K. E. (2018). Financial Incentives and Vulnerable Populations — Will Alternative Payment Models Help or Hurt? *New England Journal of Medicine*, *378*(11), 977–979. https://doi.org/10.1056/NEJMp1715455

Joynt Maddox, K. E., Reidhead, M., Hu, J., Kind, A. J. H., Zaslavsky, A. M., Nagasako, E. M., & Nerenz, D. R. (2019). Adjusting for social risk factors impacts performance and penalties in the hospital readmissions reduction program. *Health Services Research*, *54*(2), 327–336. https://doi.org/10.1111/1475-6773.13133

Joynt Maddox, K. E., Reidhead, M., Qi, A. C., & Nerenz, D. R. (2019). Association of Stratification by Dual Enrollment Status With Financial Penalties in the Hospital Readmissions Reduction Program. *JAMA Internal Medicine*, *179*(6), 769. https://doi.org/10.1001/jamainternmed.2019.0117

Kozhimannil, K. B., Hardeman, R. R., & Henning-Smith, C. (2017). Maternity care access, quality, and outcomes: A systems-level perspective on research, clinical, and policy needs. *Seminars in Perinatology*. https://doi.org/10.1053/j.semperi.2017.07.005

Kumar, A., Mahmoudi, E., & Rivera-Hernandez, M. (2019). Reducing Health Disparities in the Era of Value-Based Care. *Innovation in Aging*, *3*(Supplement_1), S562–S563. https://doi.org/10.1093/geroni/igz038.2080

Lakdawalla, D., & Philipson, T. (1998). Nonprofit production and competition. *NBER Working Paper No.6377*. Retrieved from https://www.nber.org/system/files/working_papers/w6377/w6377.pdf

Lakdawalla, D., & Philipson, T. (2006). The nonprofit sector and industry performance. *Journal of Public Economics*, *90*(8–9), 1681–1698. https://doi.org/10.1016/j.jpubeco.2005.11.004

LaVeist, T., Pollack, K., Thorpe, R., Fesahazion, R., & Gaskin, D. (2011). Place, not race: Disparities dissipate in Southwest Baltimore when blacks and whites live under similar conditions. *Health Affairs*, *30*(10), 1880–1887. https://doi.org/10.1377/hlthaff.2011.0640

Lee, J. T., Polsky, D., Fitzsimmons, R., & Werner, R. M. (2020). Proportion of Racial Minority Patients and Patients With Low Socioeconomic Status Cared for by Physician Groups After Joining Accountable Care Organizations. *JAMA Network Open*, *3*(5), e204439. https://doi.org/10.1001/jamanetworkopen.2020.4439

Leibbrand, C., Massey, C., Alexander, J. T., Genadek, K. R., & Tolnay, S. (2020). The Great Migration and Residential Segregation in American Cities during the Twentieth Century. *Social Science History*, *44*(1), 19–55. https://doi.org/10.1017/ssh.2019.46

Magee, H., Davis, L.-J., & Coulter, A. (2003). Public Views on Healthcare Performance Indicators and Patient Choice. *Journal of the Royal Society of Medicine*, *96*(7), 338–342. https://doi.org/10.1177/014107680309600707

Markovitz, A. A., & Ryan, A. M. (2017). Pay-for-Performance: Disappointing Results or Masked Heterogeneity? *Medical Care Research and Review*, *74*(1), 3–78. https://doi.org/10.1177/1077558715619282

Martin, J. R., Wang, T. Y., Loriaux, D., Desai, R., Kuchibhatla, M., Karikari, I. O., … Gottfried, O. N. (2017). Race as a predictor of postoperative hospital readmission after spine surgery. *Journal of Clinical Neuroscience*, *46*, 21–25. https://doi.org/10.1016/j.jocn.2017.08.015

Massey, D. S. (2015). The Legacy of the 1968 Fair Housing Act. *Sociological Forum*, *30*(S1), 571–588. https://doi.org/10.1111/socf.12178

McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., & Kerr, E. A. (2003). The Quality of Health Care Delivered to Adults in the United States. *New England Journal of Medicine*, *348*(26), 2635–2645. https://doi.org/10.1056/NEJMsa022615

McWilliams, J. M., Chernew, M. E., & Landon, B. E. (2017). Medicare ACO Program Savings Not Tied To Preventable Hospitalizations Or Concentrated Among High-Risk Patients. *Health Affairs*, *36*(12), 2085–2093. https://doi.org/10.1377/hlthaff.2017.0814

Meddings, J., Reichert, H., Smith, S. N., Iwashyna, T. J., Langa, K. M., Hofer, T. P., & McMahon, L. F. (2017). The Impact of Disability and Social Determinants of Health on Condition-Specific Readmissions beyond Medicare Risk Adjustments: A Cohort Study. *Journal of General Internal Medicine*, *32*(1), 71–80. https://doi.org/10.1007/s11606-016-3869-x

Min, C. J., Ehrenthal, D. B., & Strobino, D. M. (2015). Investigating racial differences in risk factors for primary cesarean delivery. *American Journal of Obstetrics and Gynecology*, *212*(6), 814.e1-814.e14. https://doi.org/10.1016/J.AJOG.2015.01.029

National Conference of State Legislatures. (2020). Maternity Length of Stay Rules. Retrieved May 11, 2020, from https://www.ncsl.org/research/health/final-maternity-length-of-stay-rules-published.aspx

National Practitioners Data Bank. (2019). The NPDB: Public Use Data File. Retrieved May 11, 2020, from https://www.npdb.hrsa.gov/resources/publicData.jsp

Newhouse, J. P. (1970). Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital. *American Economic Review*, *60*(1), 64–74. https://doi.org/10.1126/science.151.3712.867-a

Oribhabor, G. I., Nelson, M. L., Buchanan-Peart, K.-A. R., & Cancarevic, I. (2020). A Mother's Cry: A Race to Eliminate the Influence of Racial Disparities on Maternal Morbidity and Mortality Rates Among Black Women in America. *Cureus*, *12*(7). https://doi.org/10.7759/cureus.9207

Paik, M., Black, B., & Hyman, D. A. (2016). Damage caps and the labor supply of physicians: Evidence from the third reform wave. *American Law and Economics Review*, *18*(2), 463–505. https://doi.org/10.1093/aler/ahw009

Rana, R. H., Alam, K., & Gow, J. (2018). Health expenditure, child and maternal mortality nexus: a comparative global analysis. *BMC International Health and Human Rights*, *18*(1), 29. https://doi.org/10.1186/s12914-018-0167-1

Rathi, V. K. (2021, January 19). Physician Performance in the Medicare Merit-based Incentive Payment System. *JAMA - Journal of the American Medical Association*, Vol. 325, p. 308. https://doi.org/10.1001/jama.2020.22800

Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. *Demography*, *45*(3), 489–514. https://doi.org/10.1353/dem.0.0019

Roberts, E. T., Zaslavsky, A. M., Barnett, M. L., Landon, B. E., Ding, L., & McWilliams, J. M. (2018). Assessment of the Effect of Adjustment for Patient Characteristics on Hospital Readmission Rates: Implications for Pay for Performance. *JAMA Internal Medicine*, *178*(11), 1498–1507. https://doi.org/10.1001/jamainternmed.2018.4481

Rosko, M., Al-Amin, M., & Tavakoli, M. (2020). Efficiency and profitability in US not-for-profit hospitals. *International Journal of Health Economics and Management*, *20*(4), 359–379. https://doi.org/10.1007/s10754-020-09284-0

Shalowitz, D. I., Nivasch, E., Burger, R. A., & Schapira, M. M. (2018). Are patients willing to travel for better ovarian cancer care? *Gynecologic Oncology*, *148*(1), 42–48. https://doi.org/10.1016/j.ygyno.2017.10.018

Shaw, D., Guise, J.-M., Shah, N., Gemzell-Danielsson, K., Joseph, K., Levy, B., … Main, E. K. (2016). Drivers of maternity care in high-income countries: can health systems support woman-centred care? *The Lancet*, *388*(10057), 2282–2295. https://doi.org/10.1016/S0140-6736(16)31527-6

Sills, M. R., Hall, M., Cutler, G. J., Colvin, J. D., Gottlieb, L. M., Macy, M. L., … Shah, S. S. (2017). Adding Social Determinant Data Changes Children's Hospitals' Readmissions Performance. *Journal of Pediatrics*, *186*, 150-157.e1. https://doi.org/10.1016/j.jpeds.2017.03.056

Song, Z., Rose, S., Chernew, M. E., & Safran, D. G. (2017). Lower- Versus Higher-Income Populations In The Alternative Quality Contract: Improved Quality And Similar Spending. *Health Affairs*, *36*(1), 74–82. https://doi.org/10.1377/hlthaff.2016.0682

Stone, P. W. (2020, July 1). Value-Based Incentive Programs and Health Disparities. *JAMA Network Open*, Vol. 3, p. e2010231. https://doi.org/10.1001/jamanetworkopen.2020.10231

Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Author Manuscript Health Serv Outcomes Res Methodol*, *14*(4), 166–182. https://doi.org/10.1007/s10742-014-0123-z

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8 SUPPL.8). https://doi.org/10.1016/j.jclinepi.2013.01.013

Tran, L. D. (2020, June 1). Social Risk Adjustment in Health Care Performance Measures. *JAMA Network Open*, Vol. 3, p. e208020. https://doi.org/10.1001/jamanetworkopen.2020.8020

U.S. Census Bureau. (2018). *American Community Survey*. Retrieved from https://www.census.gov/programs-surveys/acs

Varkevisser, M., & Van Der Geest, S. A. (2007). Why do patients bypass the nearest hospital? An empirical analysis for orthopaedic care and neurosurgery in the Netherlands. *European Journal of Health Economics*, *8*(3), 287–295. https://doi.org/10.1007/s10198-006-0035-0

Victoor, A., Delnoij, D. M., Friele, R. D., & Rademakers, J. J. (2012). Determinants of patient choice of healthcare providers: A scoping review. *BMC Health Services Research*, Vol. 12, p. 272. https://doi.org/10.1186/1472-6963-12-272

Victoor, A., Rademakers, J., Reitsma-van Rooijen, M., De Jong, J., Delnoij, D., & Friele, R. (2014). The effect of the proximity of patients' nearest alternative hospital on their intention to search for information on hospital quality. *Journal of Health Services Research and Policy*, *19*(1), 4–11. https://doi.org/10.1177/1355819613498380

Wasserman, J., Palmer, R. C., Gomez, M. M., Berzon, R., Ibrahim, S. A., & Ayanian, J. Z. (2019). Advancing Health Services Research to Eliminate Health Care Disparities. *American Journal of Public Health*, *109*(S1), S64–S69. https://doi.org/10.2105/AJPH.2018.304922

Werbeck, A., Wübker, A., & Ziebarth, N. R. (2020). Cream Skimming by Health Care Providers and Inequality in Health Care Access: Evidence from a Randomized Field Experiment. *IZA Discussion Papers*. Retrieved from https://ideas.repec.org/p/iza/izadps/dp13100.html

Werner, R. M., Kanter, G. P., & Polsky, D. (2019). Association of Physician Group Participation in Accountable Care Organizations With Patient Social and Clinical Characteristics. *JAMA Network Open*, *2*(1), e187220. https://doi.org/10.1001/jamanetworkopen.2018.7220

Yasaitis, L. C., Pajerowski, W., Polsky, D., & Werner, R. M. (2016). Physicians' participation in ACOs is lower in places with vulnerable populations than in more affluent communities. *Health Affairs*, *35*(8), 1382–1390. https://doi.org/10.1377/hlthaff.2015.1635

Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals of Translational Medicine*, *7*(1), 16–16. https://doi.org/10.21037/atm.2018.12.10

3.9. Appendix Figures and Tables

3.9.1. Appendix Figures

Figure 3-3 Interpretation Matrix for Linear Combinations of Coefficients in Difference-in-Difference-in-Differences Analysis Comparing Effects of APII in Black Metropolitan Statistical Areas versus White Metropolitan Statistical Areas

Linear Combination	APII Implementation Period	Result	Interpretation for High-Value Outcomes	Interpretation for Low-Value Outcomes
$\beta_1 - \beta_2$	Partial (2013)	< 0	**White** MSAs experience greater improvement in quality than Black MSAs as a result of APII	**Black** MSAs experience greater improvement in quality than White MSAs as a result of APII
		> 0	**Black** MSAs experience greater improvement in quality than White MSAs as a result of APII	**White** MSAs experience greater improvement in quality than Black MSAs as a result of APII
$\beta_3 - \beta_4$	Full (2014-2016)	< 0	**White** MSAs experience greater improvement in quality than Black MSAs as a result of APII	**Black** MSAs experience greater improvement in quality than White MSAs as a result of APII
		> 0	**Black** MSAs experience greater improvement in quality than White MSAs as a result of APII	**White** MSAs experience greater improvement in quality than Black MSAs as a result of APII

Notes: Coefficients used in linear combinations are from Equation [8]. High-value outcomes refer to all prenatal and postpartum quality measures. Low-value outcomes refer to the intrapartum low-risk c-section quality measure.

3.9.2. Appendix Tables

Table 3-4 Tests for Equality in Pre-Arkansas Payment Improvement Initiative Trends in Quality of Care Outcomes: Metropolitan Statistical Areas with a High Black versus High White Population

	Prenatal Screenings						**Intrapartum Care**	**Postpartum Care**
	Linked to Gainsharing		Not Linked to Gainsharing					
	Group B Strep	Chlamydia	HIV	Hepatitis B	Gest. Diabetes	A. Bact.	Low-Risk C-Sections	Postpartum Care Within 6 Weeks
Treat * Quarter								
Black	-0.001	-0.037	0.050	0.025	0.010	-0.178**	-0.003	-0.034
	(0.056)	(0.101)	(0.046)	(0.058)	(0.103)	(0.088)	(0.051)	(0.074)
White	0.045	-0.200	-0.062	0.012	-0.158	-0.215	-0.009	0.005
	(0.062)	(0.145)	(0.148)	(0.087)	(0.299)	(0.235)	(0.051)	(0.077)
Difference	-0.045	0.163	0.113	0.013	0.168	0.036	0.006	-0.039
	(0.061)	(0.148)	(0.141)	(0.079)	(0.250)	(0.190)	(0.058)	(0.092)
N	53,630	53,630	53,630	53,630	53,630	53,630	53,630	53,630
Black MSAs								
Pre-APII								
Treatment	80.43	66.22	78.02	78.55	75.87	89.81	27.35	55.23
Control	84.24	62.61	73.52	73.67	76.01	89.32	21.09	53.06
White MSAs								
Pre-APII								
Treatment	64.06	31.50	59.62	60.47	65.96	76.32	17.76	45.67
Control	82.97	67.76	84.46	81.73	83.90	90.44	16.18	64.24

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include pre-policy treatment-quarter trend coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Black population is determined by share of Black residents from 2010-2012, prior to policy implementation being above a threshold of 12.59%, which was the corresponding national average. Covariates include all variables in Table 3-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 3-5 Covariate Balance Before and After Multiple Propensity Score Weights

	Pre-Medicaid Policy (2010-2012)						**Post-Medicaid Policy (2013-2016)**					
	Initial Balance			*Propensity Score Balance*			*Initial Balance*			*Propensity Score Balance*		
	Treatment	Control	SMD	Treatment	Control	SMD	Treatment	Control	SMD	Treatment	Control	SMD
Maternal Characteristics												
Maternal Age 35+	0.09	0.14	-0.15	0.09	0.09	0.01	0.09	0.14	-0.15	0.09	0.09	-0.01
LOS > 4	0.01	0.03	-0.11	0.01	0.01	0.02	0.02	0.03	-0.10	0.02	0.02	0.00
Health Plan												
HMO	0.02	0.30	-0.84	0.02	0.02	-0.02	0.01	0.19	-0.60	0.01	0.01	0.00
PPO	0.77	0.48	0.62	0.77	0.74	0.06	0.68	0.49	0.39	0.68	0.68	-0.02
POS	0.09	0.10	-0.04	0.09	0.11	-0.09	0.09	0.06	0.12	0.09	0.09	0.02
High Deductible	0.13	0.12	0.03	0.13	0.13	0.01	0.22	0.27	-0.12	0.22	0.21	0.01
% Cost Sharing	0.64	0.45	0.39	0.64	0.66	-0.04	0.79	0.54	0.54	0.79	0.79	0.00
MSA-Level Healthcare Characteristics												
% Hospitals Non-Profit	0.05	0.51	-1.19	0.05	0.08	-0.08	0.26	0.51	-0.53	0.26	0.28	-0.04
% Hospitals Provide OB Services	0.13	0.46	-0.80	0.13	0.19	-0.15	0.12	0.40	-0.67	0.12	0.14	-0.04
% Teaching Hospitals	0.38	0.59	-0.44	0.38	0.44	-0.12	0.49	0.52	-0.06	0.49	0.57	-0.16
# FQHCs	0.46	0.86	-0.92	0.46	0.43	0.08	0.76	0.88	-0.30	0.76	0.76	0.01
Beds per 1,000	1.00	0.45	1.56	1.00	0.91	0.25	1.00	0.47	1.50	1.00	0.79	0.58
PCPs per 1,000	0.50	0.45	0.11	0.50	0.54	-0.08	0.64	0.51	0.26	0.64	0.61	0.06
MSA-Level Demographic Characteristics												
% Less Than HS Education	0.65	0.75	-0.22	0.65	0.76	-0.23	0.48	0.46	0.04	0.48	0.49	-0.03
% More Than College Education	0.00	0.42	-1.20	0.00	0.26	-0.75	0.38	0.64	-0.52	0.38	0.48	-0.20
MSA-Level Economic Characteristics												
% Uninsured	1.00	0.77	0.78	1.00	0.89	0.36	0.38	0.48	-0.20	0.38	0.63	-0.50
% Unemployed	0.38	0.92	-1.38	0.38	0.52	-0.37	0.19	0.27	-0.19	0.19	0.17	0.06
% Poverty	0.62	0.57	0.11	0.62	0.71	-0.18	0.43	0.38	0.11	0.43	0.45	-0.03
% Population Black	0.44	0.47	-0.06	0.44	0.50	-0.11	0.58	0.54	0.08	0.58	0.58	0.01

Notes: Propensity scores constructed using logistic regression. Covariates include binary indicators for each variable, coded as 1 if the value in a given MSA-Quarter is ≥ median and 0 if < median. Select covariates are excluded due to potential influence from the Medicaid payment policy in the post-period, including Medicaid share and OB malpractice payout. The model also included interaction terms for each of the covariates listed in Table 3-5 with an indicator for High Black MSA, to increase covariate balance between High White and High Black MSAs. The estimate of interest is the Standardized Mean Difference (SMD), which provides an independent comparison between treated and control means.

Table 3-6 Sensitivity Analysis - Variation in Effects of the Arkansas Payment Improvement Initiative Between Metropolitan Statistical Areas with a High Black Population versus High White Population: Unweighted Analysis

	Prenatal Screenings: Linked to Gainsharing						**Intrapartum Care**	
	Group B Strep		Chlamydia		HIV		Low-Risk C-Section	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post								
Black	-0.204	0.250	0.371***	0.469***	-0.234*	0.468**	0.091*	-0.169
	(0.275)	(0.262)	(0.142)	(0.170)	(0.129)	(0.204)	(0.051)	(0.117)
White	0.535**	0.740***	0.976***	1.164***	0.598***	0.535**	-0.260	0.111
	(0.224)	(0.263)	(0.218)	(0.302)	(0.164)	(0.261)	(0.187)	(0.100)
Difference	-0.738**	-0.490	-0.605**	-0.695**	-0.833***	-0.067	0.351*	-0.280**
	(0.348)	(0.359)	(0.249)	(0.328)	(0.155)	(0.296)	(0.187)	(0.137)
N	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858

	Prenatal Screenings: Not Linked to Gainsharing						**Postpartum Care**	
	Hepatitis B		Gestational Diabetes		Asympt. Bact.		Postpartum Visit within 6 Weeks	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post								
Black	-0.350**	0.140	0.109	0.321	-0.448***	0.408	0.177**	0.344***
	(0.150)	(0.195)	(0.254)	(0.274)	(0.160)	(0.312)	(0.088)	(0.096)
White	0.368***	0.063	0.305	0.596***	0.490**	0.731***	0.467**	0.532***
	(0.126)	(0.166)	(0.227)	(0.209)	(0.215)	(0.278)	(0.185)	(0.202)
Difference	-0.718***	0.077	-0.196	-0.275	-0.938***	-0.323	-0.290	-0.188
	(0.182)	(0.256)	(0.330)	(0.339)	(0.234)	(0.401)	(0.200)	(0.212)
N	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Black population is determined by share of Black residents from 2010-2012, prior to policy implementation being above a threshold of 12.59%, which was the corresponding national average. Covariates include all variables in Table 3-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table 3-7 Sensitivity Analysis - Variation in Effects of the Arkansas Payment Improvement Initiative Between Metropolitan Statistical Areas with a High Black Population versus High White Population: With Excluded Episodes

	Prenatal Screenings: Linked to Gainsharing						**Intrapartum Care**	
	Group B Strep		Chlamydia		HIV		Low-Risk C-Section	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post								
Black	0.048	-0.343	0.092	0.250	-0.213	-0.097	0.015	-0.151
	(0.203)	(0.306)	(0.223)	(0.217)	(0.192)	(0.273)	(0.104)	(0.153)
White	0.005	0.348	0.560*	0.689**	0.188	-0.043	-0.218	0.243*
	(0.173)	(0.215)	(0.327)	(0.297)	(0.215)	(0.312)	(0.139)	(0.143)
Difference	0.042	-0.691**	-0.468	-0.439	-0.400	-0.054	0.232*	-0.394**
	(0.277)	(0.273)	(0.335)	(0.304)	(0.288)	(0.332)	(0.141)	(0.175)
N	227,817	227,817	227,817	227,817	227,817	227,817	227,817	227,817

	Prenatal Screenings: Not Linked to Gainsharing						**Postpartum Care**	
	Hepatitis B		Gestational Diabetes		Asympt. Bact.		Postpartum Visit within 6 Weeks	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post								
Black	-0.214	-0.340	0.182	-0.824**	-0.163	-0.256	-0.163	0.263
	(0.172)	(0.252)	(0.173)	(0.342)	(0.247)	(0.284)	(0.132)	(0.178)
White	0.329	-0.322	0.698***	0.697**	0.147	0.111	0.339**	0.439**
	(0.246)	(0.283)	(0.204)	(0.330)	(0.388)	(0.213)	(0.147)	(0.208)
Difference	-0.543*	-0.018	-0.516*	-1.521***	-0.310	-0.368	-0.604**	-0.225
	(0.327)	(0.309)	(0.277)	(0.359)	(0.358)	(0.268)	(0.273)	(0.258)
N	227,817	227,817	227,817	227,817	227,817	227,817	227,817	227,817

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Black population is determined by share of Black residents from 2010-2012, prior to policy implementation being above a threshold of 12.59%, which was the corresponding national average. Covariates include all variables in Table 3-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 3-8 Sensitivity Analysis - Variation in Effects of the Arkansas Payment Improvement Initiative Between Metropolitan Statistical Areas with a High Black Population versus High White Population: Testing the Percent of Population Black as a Continuous Variable

	Prenatal Screenings: Linked to Gainsharing						**Intrapartum Care**	
	Group B Strep		Chlamydia		HIV		Low-Risk C-Section	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * % Black	-0.021*	-0.038***	-0.026**	-0.027**	-0.010	-0.004	0.009	-0.019**
	(0.011)	(0.010)	(0.013)	(0.013)	(0.010)	(0.019)	(0.008)	(0.007)
N	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858

	Prenatal Screenings: Not Linked to Gainsharing						**Postpartum Care**	
	Hepatitis B		Gestational Diabetes		Asympt. Bact.		Postpartum Visit within 6 Weeks	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * % Black	-0.017	0.000	-0.015	-0.049**	-0.010	-0.016	-0.010	0.013
	(0.014)	(0.000)	(0.011)	(0.020)	(0.017)	(0.012)	(0.010)	(0.012)
N	158,858	158,858	158,858	158,858	158,858	158,858	158,858	158,858

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Black population is determined by share of Black residents from 2010-2012, prior to policy implementation being above a threshold of 12.59%, which was the corresponding national average. Covariates include all variables in Table 3-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 3-9 Sensitivity Analysis - Variation in Effects of the Arkansas Payment Improvement Initiative Between Metropolitan Statistical Areas with a High Black Population versus High White Population: Alternate Control Groups

	Prenatal Screenings: Linked to Gainsharing						**Intrapartum Care**	
	Group B Strep		Chlamydia		HIV		Low-Risk C-Section	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * [Black – White]								
No CA	-0.393	-0.821***	-0.360	-0.331	-0.597**	-0.120	0.418*	-0.248
	(0.310)	(0.294)	(0.356)	(0.347)	(0.288)	(0.320)	(0.250)	(0.200)
No CT	-0.264	-0.787***	-0.364	-0.364	-0.519*	-0.089	0.378	-0.245
	(0.330)	(0.298)	(0.361)	(0.343)	(0.303)	(0.310)	(0.229)	(0.196)
No FL	-0.201	-0.643*	-0.460	-0.307	-0.615**	-0.069	0.530*	-0.196
	(0.337)	(0.335)	(0.376)	(0.328)	(0.301)	(0.325)	(0.273)	(0.226)
No GA	-0.537*	-0.943***	-0.462	-0.423	-0.680***	0.146	0.356	-0.306
	(0.295)	(0.247)	(0.341)	(0.354)	(0.254)	(0.317)	(0.275)	(0.204)
No IN	-0.402	-0.810***	-0.358	-0.288	-0.641**	-0.124	0.488**	-0.198
	(0.298)	(0.297)	(0.320)	(0.332)	(0.272)	(0.323)	(0.249)	(0.207)
No KY	-0.338	-0.732**	-0.443	-0.450	-0.584**	-0.148	-0.490*	-0.203
	(0.311)	(0.326)	(0.342)	(0.345)	(0.282)	(0.320)	(0.258)	(0.214)
No MI	-0.585**	-1.020***	-0.564*	-0.523	-0.663**	-0.201	0.519**	-0.168
	(0.286)	(0.250)	(0.329)	(0.337)	(0.294)	(0.339)	(0.259)	(0.213)
No MO	-0.340	-0.773***	-0.428	-0.378	-0.622**	-0.127	0.484**	-0.202
	(0.307)	(0.285)	(0.339)	(0.338)	(0.274)	(0.311)	(0.239)	(0.196)
No NJ	-0.340	-0.774***	-0.421	-0.372	-0.621**	-0.127	0.485**	-0.202
	(0.306)	(0.286)	(0.338)	(0.337)	(0.274)	(0.310)	(0.239)	(0.197)
No TX	-0.286	-0.511	-0.482	-0.445	-0.757***	-0.106	0.484***	-0.104
	(0.312)	(0.340)	(0.404)	(0.455)	(0.260)	(0.308)	(0.186)	(0.182)
No VA	-0.089	-0.743**	-0.119	-0.408	-0.443	-0.392	0.679***	-0.082
	(0.330)	(0.320)	(0.360)	(0.354)	(0.363)	(0.294)	(0.223)	(0.192)
No WI	-0.312	-0.738**	-0.417	-0.359	-0.608**	-0.130	0.490**	-0.183
	(0.313)	(0.289)	(0.343)	(0.346)	(0.277)	(0.324)	(0.243)	(0.197)

	Prenatal Screenings: Not Linked to Gainsharing						**Postpartum Care**	
	Hepatitis B		Gestational Diabetes		Asympt. Bact.		Postpartum Visit 6 Weeks	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * [Black – White]								
No CA	-0.580*	0.095	-0.619*	-1.507***	-0.516	-0.412	-0.651**	-0.244
	(0.323)	(0.257)	(0.322)	(0.415)	(0.336)	(0.279)	(0.283)	(0.261)
No CT	-0.522	0.072	-0.595*	-1.521***	-0.498	-0.534*	-0.612**	-0.240
	(0.335)	(0.272)	(0.320)	(0.392)	(0.351)	(0.293)	(0.279)	(0.263)
No FL	-0.548	0.234	-0.573*	-1.474***	-0.522	-0.373	-0.539*	-0.173
	(0.365)	(0.258)	(0.316)	(0.417)	(0.400)	(0.321)	(0.278)	(0.237)
No GA	-0.637**	0.137	-0.569*	-1.424***	-0.737**	-0.445	-0.425**	-0.159
	(0.261)	(0.272)	(0.334)	(0.384)	(0.328)	(0.302)	(0.201)	(0.212)
No IN	-0.597**	0.057	-0.588*	-1.462***	-0.546*	-0.451	-0.594**	-0.239
	(0.291)	(0.259)	(0.302)	(0.397)	(0.329)	(0.309)	(0.274)	(0.254)
No KY	-0.606**	-0.041	-0.555*	-1.534***	-0.640**	-0.507	-0.615**	-0.237
	(0.298)	(0.270)	(0.312)	(0.395)	(0.310)	(0.312)	(0.280)	(0.271)

No MI	-0.726**	-0.080	-0.587*	-1.482***	-0.685*	-0.575*	-0.580**	-0.212
	(0.311)	(0.267)	(0.322)	(0.385)	(0.358)	(0.303)	(0.277)	(0.273)
No MO	-0.587**	0.069	-0.557*	-1.450***	-0.573*	-0.470	-0.604**	-0.225
	(0.297)	(0.262)	(0.308)	(0.389)	(0.324)	(0.291)	(0.273)	(0.258)
No NJ	-0.589**	0.058	-0.559*	-1.456***	-0.585*	-0.492*	-0.597**	-0.212
	(0.297)	(0.259)	(0.307)	(0.390)	(0.323)	(0.291)	(0.276)	(0.259)
No TX	-0.625***	-0.070	-0.261	-0.784***	-0.589**	-0.467	-0.525*	-0.200
	(0.216)	(0.299)	(0.271)	(0.258)	(0.244)	(0.357)	(0.276)	(0.247)
No VA	-0.327	0.144	-0.457	-1.544***	-0.424	-0.577*	-0.727**	-0.317
	(0.341)	(0.290)	(0.324)	(0.389)	(0.359)	(0.306)	(0.316)	(0.281)
No WI	-0.569*	0.088	-0.527*	-1.418***	-0.523	-0.407	-0.637**	-0.276
	(0.301)	(0.268)	(0.311)	(0.387)	(0.330)	(0.295)	(0.280)	(0.265)

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Black population is determined by share of Black residents from 2010-2012, prior to policy implementation being above a threshold of 12.59%, which was the corresponding national average. Covariates include all variables in Table 3-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

4. Geographic Variation in Effects of Mandatory Bundled Payments by Baseline Performance

4.1. Introduction

Since the Affordable Care Act (ACA), provider reimbursement has moved away from fee-for-service (FFS) towards payment models that incentivize value. One key approach has been bundling payments for specific clinical episodes, which shifts financial risk to physicians by offering a single, risk-adjusted case rate for the entire course of care, rather than reimbursing separately for each individual service (C. Carroll et al., 2018). Under this policy design, providers that keep costs below a risk-adjusted target threshold and meet certain benchmarks on quality measures earn a portion of the savings, while those that exceed it incur a financial penalty. Holding providers accountable for quality and costs of care during the episode creates a financial incentive to coordinate care and improve the patient experience across both dimensions (Agarwal, Liao, Gupta, & Navathe, 2020).

The Centers for Medicare and Medicaid Services (CMS) has promoted bundled payments as a flagship policy approach for value-based payment reform. Medicare launched three major bundled payment programs in 2009, 2013, and 2016, respectively. The Acute Care Episode (ACE) Demonstration and Bundled Payments for Care Improvement (BPCI) are voluntary initiatives that cover all Medicare charges for hospitalizations and post-acute care (PAC) across select surgical and cardiac clinical episodes; and the Comprehensive Care for Joint Replacement (CJR) model is a mandatory program for hip and knee replacements, in which hospitals are responsible for all Medicare inpatient spending plus 90 days post-discharge. CJR hospitals were selected based on location within a high-spending region (Agarwal et al., 2020). With the exception of CJR, these programs rely on voluntary participation, which can introduce selection issues if participating hospitals differ from non-participating hospitals. Specifically, if unobservable characteristics are driving a hospital's choice to participate in a bundled payment program, then estimates of the program effects may be biased by factors that cannot be captured in a dataset. As a result, the impact of bundled payments gleaned from current studies may differ from the general provider population, and pathways for achieving success may be misleading. This poses a challenge for policymakers, who may not be fully informed as to whether bundled payments can work in today's healthcare landscape (Glickman, Dinh, & Navathe, 2018).

Empirical studies have found that federal bundled payment programs have led to modest reductions in spending (Barnett et al., 2019; Curtin, Russell, & Odum, 2017; Dummit et al., 2016; Navathe et al., 2018, 2017; Odum, Hamid, Van Doren, & Spector, 2018; Siddiqi et al., 2018), but effects on quality are mixed (Barnett et al., 2019; Bhatt et al., 2017; Bolz & Iorio, 2016; Bronson et al., 2019; Chen, Ryan, Shih, Thumma, & Dimick, 2018; Courtney, Ashley, Hume, & Kamath, 2016; Curtin et al., 2017; Jubelt et al., 2017; Kee, Edwards, & Barnes, 2017; B. I. Martin, Lurie, Farrokhi, McGuire, & Mirza, 2018; Odum et al., 2018; Siddiqi et al., 2018). However, most existing work focuses on the impact of bundled payments for the "average" self-selected participant, but little is known about how effects vary across providers. Further, results are specific to the Medicare population, leading to concerns that findings cannot be generalized to younger, non-Medicare populations. BPCI-participating hospitals are more likely to be large, high-volume, non-profit, teaching facilities with integrated PAC facilities (Navathe et al., 2018). Further, within this self-selected group, the greatest spending reductions are concentrated among top performers at baseline (Navathe et al., 2017). While these successes are important, they do not clearly indicate whether bundled payments address spending and quality among providers facing the biggest challenges, such as low quality, a high-risk patient population, and limited resources.

To shed light on the effects of bundled payments across all providers, it is critical to assess the impact of episode-based payments with mandatory, rather than voluntary participation. Evaluating mandatory bundled payments to explore whether provider behavior varies by baseline levels of quality is key to determining the net social welfare effects of these incentives (e.g. maximizing benefits and minimizing harms). The answer to this question is not straightforward;

providers with lower quality at baseline face a higher marginal cost to change behavior, but they also have greater room for improvement.

In this study, I test whether baseline provider quality is associated with heterogeneity in effects on quality and spending outcomes of a mandatory bundled payment program. It is important to explore variation in these effects, since the "average" effect may attenuate towards the null if providers with different baseline characteristics experience divergent effects (Markovitz & Ryan, 2017). To achieve the study aims, I leverage a unique, multi-payer, state-level episode-based payment program for perinatal care, with compulsory provider participation for Medicaid and commercially insured enrollees. This policy is distinct from the conventional bundled payment structure because participation was required across all providers and targeted younger patients insured by Medicaid and private payers (C. Carroll et al., 2018). Further, this context allows me to decompose the drivers of episode-based cost savings by quantity and price, to determine whether the mechanism for cost savings varies between low and high quality providers at baseline. I use a difference-in-differences design and a large commercial claims database. Overall, this study seeks to inform whether the effects of mandatory versus voluntary payment reform are expected to vary, to offer insight into future state and federal bundled payment design.

4.2. Study Data and Methods

4.2.1. Arkansas Payment Improvement Initiative Overview

The Arkansas Payment Improvement Initiative (APII) is a statewide multi-payer episode-based payment program with compulsory provider participation. APII was introduced in 2013 for select episodes in Medicaid and private insurers, including perinatal care. Under APII, insurers set episode-level spending targets and hold a Principal Accountable Provider (PAP) responsible for a portion of excess spending for perinatal episodes. Specifically, PAPs are held responsible for half of total risk-adjusted episode spending, either losing half of the excess spending or earning half of the savings. Further, PAPs are only eligible to receive payment conditional upon achieving 80%

screening rates for the following prenatal process measures: Human Immunodeficiency Virus (HIV), Group B Streptococcus (Group B Strep), and chlamydia. The program was implemented in two distinct stages: (1) in 2013, Medicaid implemented APII across nearly all beneficiaries, and (2) in 2014, participating private insurers launched APII among self-insured and fully-insured groups. Due to its stepwise rollout, I assess effects of APII in two separate time periods: (1) after partial implementation in 2013 (only Medicaid enrollees were covered by APII), and (2) after full implementation in 2014 (Medicaid and commercial enrollees were both covered by APII). Previous research examining APII has found a decline in average perinatal spending in Arkansas by a significant 3.8% compared to surrounding states, with a small but insignificant increase in quality of care. Researchers attributed savings to reduced inpatient prices, as opposed to significant declines in per episode quantity of care (C. Carroll et al., 2018).

4.2.2. Effects of Medicare Bundled Payments on Spending and Quality

Empirical studies on Medicare bundled payments provide evidence on the association between bundled payments and episode savings, with greater spending reductions in voluntary programs relative to mandatory ones. Five studies showed a significant decrease in episode spending for lower extremity joint replacements among BPCI participants (Curtin et al., 2017; Dummit et al., 2016; Navathe et al., 2017; Odum et al., 2018; Siddiqi et al., 2018). One multi-center analysis found a 4% decline in per episode spending ($1,166) among BPCI participants compared to non-participants, with the largest savings concentrated in the post-acute care (PAC) setting (Dummit et al., 2016). Three single-center studies also showed significant reductions in episode spending among BPCI participants, ranging from $1,717 to $3,263 per episode (Curtin et al., 2017; Siddiqi et al., 2018). Savings were primarily driven by reductions in length of stay, PAC use, and internal hospital costs, and were limited to uncomplicated episodes (Navathe et al., 2017). Current work suggests that savings did not extend to the ACE program or other types of clinical episodes, so generalizability of BPCI results for joint replacements may be limited (Joynt Maddox, Orav, Zheng, & Epstein, 2018; Navathe et al., 2018, 2017). Research on CJR is suggestive of episode savings, but since the program is relatively new, findings are less conclusive. One study on CJR demonstrated a significant 3.6% reduction in spending ($1,084) (Barnett et al., 2019). Another study comparing BPCI with CJR determined $234 greater savings per episode in BPCI, but the difference was not significant (Finkelstein et al., 2018). Like BPCI, spending reductions in CJR were driven by lower PAC expenditures (Barnett et al., 2019; Finkelstein et al., 2018).

The association between bundled payments and quality improvement are mixed. Six studies showed significant reductions in readmission rates, ranging from 0.6% to 7.0% (Barnett et al., 2019; Chen et al., 2018; Curtin et al., 2017; Jubelt et al., 2017; Odum et al., 2018; Siddiqi et al., 2018), while eleven studies did not (Bhatt et al., 2017; Bolz & Iorio, 2016; Bronson et al., 2019; Courtney et al., 2016; Dummit et al., 2016; Finkelstein et al., 2018; Gray, Prieto, Duncan, & Parvataneni, 2018; Joynt Maddox et al., 2018; Kee et al., 2017; B. I. Martin et al., 2018; Navathe et al., 2017). Changes on other dimensions of quality, including complication rates (Barnett et al., 2019; Chen et al., 2018; Finkelstein et al., 2018; Gray et al., 2018), mortality (Barnett et al., 2019; Chen et al., 2018; Dummit et al., 2016; Joynt Maddox et al., 2018), and emergency department visits (Barnett et al., 2019; Dummit et al., 2016; Finkelstein et al., 2018; Joynt Maddox et al., 2018; Navathe et al., 2017), have not been observed.

I extend this literature by exploring variation in the effects of mandatory bundled payments by providers' baseline performance. Assessing such heterogeneity seeks to establish whether the "average" effects that have been explored to date mask divergent effects within the provider population. This question is critical to assessing whether episode-based payments with mandatory, rather than voluntary participation require different approaches across providers with different baseline characteristics, which is critically important in determining the net social welfare effects.

4.2.3. Commercial Claims Data

The primary data for this analysis is the Truven MarketScan Commercial Claims database from 2010 to 2016, which links paid claims and encounter data with detailed patient information across sites and types of providers over time. Although the database is a convenience sample of enrollees in commercial health plans and large self-insured firms that opt to provide their data, the MarketScan data includes proprietary commercial claims (employer and health plan) from over 36 million patient hospital discharges (Johns Hopkins, 2016). These data are collected across broad geographic areas to represent treatment patterns and costs in the U.S. The MarketScan data are advantageous for this analysis because they use consistent enrollee identifiers over time, enabling us to track patients across the full episode of care. One limitation is that a major insurer dropped out of the MarketScan data in 2015. To avoid differential selection into the database over time, the sample is limited to the employer population, which remains stable over the study period. Relying on the employer population also aims to ensure that the sample is minimally affected by changes in the insurance policy landscape, such as Medicaid expansion or the private option in Arkansas.

4.2.4. Sample

The analytic sample consists of 119,309 perinatal episodes between 2010 and 2016. This includes 2,031 episodes for providers in Arkansas and 117,278 episodes in control states. I use the MarketScan database to construct perinatal episodes using the methodology outlined in the Arkansas BCBS Perinatal Algorithm. First, I identify all live births between 2010 and 2016 using the relevant Diagnosis Related Group (DRG) codes, and pull all inpatient and outpatient claims in the period 40 weeks before the delivery through sixty days afterwards. I then collapse the data to the episode-level, using the date of birth for assignment to the pre- and post- periods (Arkansas Blue Cross Blue Shield, 2014).

The sample includes low-risk births in the Truven MarketScan Database, defined by the Agency for Healthcare Quality and Research (AHRQ). This population is limited to mothers with uncomplicated births (e.g. no abnormal presentation, preterm delivery, fetal death, multiple gestation, or breech procedure) who have never had a prior c-section. Following Carroll et al. (2018), I exclude cases that are exempt from APII reimbursement due to patient co-morbidities (e.g. sickle cell anemia, end stage renal disease, severe preeclampsia). Focusing on this population guarantees that all physicians in the treatment group face the same financial incentive, and that all quality measures are appropriate for this population (C. Carroll et al., 2018).

4.2.5. Outcome Variables

The primary outcome measure is total episode spending. Consistent with Carroll et al. (2018), I define this as the sum of all payments during the episode window. Further, I examine spending subcategories, including: intrapartum inpatient facility, intrapartum inpatient professional, total prenatal, and total postpartum. Intrapartum facility includes reimbursement to the hospital where the childbirth took place. Variation in intrapartum facility spending is driven by hospital prices, as well as the mode of delivery (e.g. because DRG payments are higher for c-sections than vaginal deliveries). Intrapartum professional spending includes the global maternity fee, which is a one-time billing for routine prenatal, intrapartum, and postpartum care. The global maternity fee is often triggered when the physician delivers the baby, and it can vary by delivery method and across physician groups that negotiate different fee schedules. Total prenatal spending encompasses all payments prior to the birth outside of the global maternity fee (e.g. laboratory tests, inpatient stays, or emergency department visits). Accordingly, total postpartum spending refers to all payments after the birth that occur outside of the global maternity fee.

I also construct eight quality outcome measures for each episode, all of which are adapted from the Maternity Care Performance Measure Set developed jointly by the American College of Obstetricians and Gynecologists and the National Committee for Quality Assurance. Quality outcomes consist of the following prenatal measures: whether a patient received three screenings linked to gainsharing (Group B Strep, Chlamydia, and HIV) and three screenings not linked to gainsharing (Hepatitis B, Gestational Diabetes, and Asymptomatic Bacteriuria). PAPs must report each of the prenatal screening rates to track their performance, regardless of whether it is tied to reimbursement. In the postpartum period, examine whether the mother received any follow-up care within eight weeks of the delivery. All prenatal and postpartum outcomes are considered high-value components of the perinatal episode, which means that a higher rate is considered better. In the intrapartum period, I measure low-risk c-sections, which identifies whether the mother received an inappropriate c-section. A c-section is not appropriate if the mother has a singleton, uncomplicated birth, and has never had a c-section before. I follow the methodology developed by AHRQ, using Inpatient Quality Indicator (IQI) 33, which is endorsed by the National Quality Forum as a consensus standard for hospital care (Agency for Healthcare Quality and Research, 2016). Unlike the other outcomes, low-risk c-sections capture a dimension of low-value care, meaning that a lower rate is considered better.

4.2.6. Covariates

Covariates include the following maternal characteristics: insurance type (HMO, PPO, POS, and high-deductible health plan), age category (less than 25, 25-29, 30-34, and 35+), whether the hospital length of stay was greater than four days (the number of days typically covered by insurers), and cost sharing quartile bins. I also included a series of MSA-level controls, which account for healthcare, demographic, and economic variables that may influence physician practice patterns. Healthcare factors consisted of hospital characteristics (percent of hospitals that are non-profit, percent of hospitals that provide obstetric services, percent teaching hospitals, number of federally qualified health centers (FQHCs), beds per 1,000, and percent of patients that are insured by Medicaid), practitioner information (primary care practitioners per 1,000), and malpractice risk quartile bins, defined as the average obstetric-related malpractice payout. Demographic characteristics referred to the percent of the population with less than a high school education and percent of the population with more than a college education. Finally, economic characteristics included uninsurance rate, percent unemployed, and percent with income below the federal poverty line. Covariates were selected to be consistent with Carroll et al., (2018). They were derived from several data sources, including the U.S. Census American Community Survey (ACS), the Health Resources and Services Administration's Area Health Resource File (AHRF), the American Hospital Association (AHA) Annual Survey, and the National Practitioners Data Bank (NPDB) (AHA, 2018; HRSA, 2019; NPDB, 2019; U.S. Census Bureau, 2018).

4.2.7. High Performance Metropolitan Statistical Areas

I assess whether the effects of APII on spending and quality differ between providers located in high performing geographic areas, compared to low performing geographic areas, measured at the metropolitan statistical area (MSA) level. To determine whether an MSA is high or low performing, I created a quality performance score, derived as an equally weighted composite of seven perinatal quality measures in the study period prior to APII implementation (2010-2012): low-risk c-sections, prenatal screenings (Group B. Strep, Asymptomatic Bacteriuria, Hepatitis B, Gestational Diabetes, HIV[1]), and postpartum follow-up within 8 weeks. Six of the seven measures are considered high-value, in which expected benefits outweigh expected harms, and thus, a higher rate is considered better. One measure, low-risk c-section rate, is a measure of low-value care, in which expected harms outweigh expected benefits, so a lower rate is considered better. For consistency in scoring, I invert the low-risk c-section measure by subtracting the rate from 100%. An MSA is assigned as high performing if it has a composite score greater than or equal to the sample mean score of 68.68%.

4.2.8. Control Group

I constructed the control group by adapting Stuart et al.'s (2014) multiple propensity score weighting approach. This method aimed to ensure that control states were similar to treatment states across observable covariates (Stuart et al., 2014). Absent this approach, selection into treatment and control groups could be confounded by baseline characteristics, thereby biasing estimates. First, I fit separate multinomial logistic regression models in the pre- and post-APII period, to predict the probability of being assigned to the treatment group in each time frame. The models adjusted for all covariates included in the main model, except for those that could potentially be

[1] Chlamydia screenings were not included in the quality performance composite score due to low rates at baseline, leading to a small sample size in the "high performing" group. In the study sample, the rate of chlamydia screenings in the pre-policy period was 46.81% in the treatment group and 54.43% in the control group. The APII quality threshold goal rate is 80%.

influenced by APII in the post-period. Second, I divided the predicted values of each logistic regression by its inverse to generate the propensity score weight. The resulting weights helped to generate a synthetic sample in which treatment assignment was independent from measured baseline covariates. The control group consisted of fifteen states that did not implement a bundled payment policy (California, Connecticut, Indiana, Kentucky, Louisiana, Michigan, Montana, North Carolina, North Dakota, New York, Pennsylvania, South Carolina, Texas, West Virginia, and Wyoming). This group represents a diverse set of states across multiple regions, which may increase external validity.

4.2.9. Statistical Analyses

I first sought to understand whether spending and quality differed between episodes in high-performing vs. low-performing MSAs upon implementation of APII. To answer this question, I used a difference-in-difference-in-differences (DDD) model comparing high and low performing MSAs in treatment and control states, pre-policy implementation (2010-2012) versus post-policy implementation (2013-2016). In the main DDD model, I estimated separate effects of APII after partial implementation (2013), when the program had mainly been rolled out in Medicaid, and full implementation (2014-2016), when the policy had also been initiated among commercial payers. This approach enabled me to assess whether there were spillovers from Medicaid onto commercially insured enrollees, and is modeled on earlier work comparing average effects of APII on spending utilization (C. Carroll et al., 2018).

I estimated the impact of APII for high and low performing MSAs using the following equation:

$$Y_{imst} = \beta_0 + \beta_1 \cdot \text{Treat}_s \cdot \text{Partial}_t \cdot \text{High}_m + \beta_2 \cdot \text{Treat}_s \cdot \text{Partial}_t \cdot \text{Low}_m + \beta_3 \cdot \text{Treat}_s \cdot \text{Full}_t \cdot \text{High}_m + \beta_4 \cdot \text{Treat}_s \cdot \text{Full}_t \cdot \text{Low}_m + \beta_5 \cdot \text{Treat}_s + \beta_6 \cdot \text{Full}_t + \beta_7 \cdot \text{Full}_t + \beta_8 \cdot \text{Treat}_s \cdot \text{Partial}_t + \beta_9 \cdot \text{High}_m \cdot \text{Partial}_t + \beta_{10} \cdot \text{Treat}_s \cdot \text{Full}_t + \beta_{11} \cdot \text{High}_m \cdot \text{Full}_t + \beta_{12} \cdot \text{High}_m \cdot \text{Treat}_s + \zeta \cdot P_i + T_t + \vartheta \cdot Z_m + \mu \cdot V_s + \varepsilon_{imst} \quad [9]$$

In [9], Y_{imst} is the expected value of the outcome. It is indexed by the episode i, State-MSA m; in state s; at time t, which is representative of pre/post policy implementation. Treat is a binary variable that denotes the presence of APII, while Partial and Full are binary variables that indicate the post-periods, including partial implementation in 2013, or full implementation in 2014-2016. High_m and Low_m, defined earlier, refer to high and low performing MSAs, respectively. V_s and T_t are state and quarter-year fixed-effects, respectively. P_i is a vector of episode-level maternal characteristics. Z_m is a vector of time-varying State-MSA-level controls, which account for healthcare, demographic, and economic variables that may influence physician practice patterns. To account for covariance in standard errors between time periods by geographic areas, standard errors are clustered at the State-MSA level. All models of healthcare spending are estimated using a one-part generalized linear model (GLM) with a Gamma distributional family and log link. All models of perinatal quality are estimated using logistic regression.

In the main model, the coefficients of interest are $\beta_1 - \beta_2$ and $\beta_3 - \beta_4$, which respectively, represent the aggregate effect of APII in high versus low performing MSAs after partial and full implementation. I calculated these estimates using a linear combination. For spending variables and low-value quality measures (e.g. low-risk c-sections), a negative value suggests that high quality MSAs experienced greater savings and/or quality gains. For high-value quality outcomes, including prenatal and postpartum measures, a positive value indicates that high quality MSAs experienced greater quality gains than low quality MSAs as a result of APII.

After examining overall savings, I performed analyses to decompose whether intrapartum savings were achieved through a reduction in prices and/or quantity of care supplied. To estimate price changes, I modified [9] by adding a fixed effect for each DRG. To estimate quantity changes, I modified equation [9] by adding a fixed effect for standardized intrapartum facility claims, using the median DRG payment. This approach was adapted from Carroll et al., (2018), and sought to determine whether holding intrapartum utilization or prices constant affected the main result (C.

Carroll et al., 2018). These analyses enabled me to explore whether the mechanism for savings varied between high and low performing MSAs.

I also performed several additional tests. First, I repeated the analyses without multiple group propensity score weights. Second, I re-ran analyses using alternate comparison groups by dropping one control state at a time. Third, I tested whether results were sensitive to inclusion of non-metropolitan areas. Fourth, I re-estimated the models using a continuous measure for composite performance. Finally, I re-ran analyses with episodes that are excluded from APII reimbursement.

4.3. Study Results

4.3.1. Descriptive Statistics

Overall, the analyses examined 119,309 low-risk pregnancy episodes. This included 2,031 episodes for providers in Arkansas and 117,278 episodes in the control group. In the treatment group, there were 1,131 episodes in high performing MSAs and 900 in low performing MSAs. In the control group, there were 104,622 episodes in high performing MSAs and 12,656 in low performing MSAs.

Table 4-1 summarizes sample characteristics in high and low performing MSAs in treatment and control states, before and after implementation of APII. Differences between groups were relatively small in magnitude during the pre-policy period, when comparing low performing and high performing MSAs, respectively, in the treatment group versus the control group. This was true for episode-level maternal characteristics, and MSA-level healthcare and economic factors. Prior to implementing APII, high performing MSAs in both treatment and control groups had a higher proportion of HMO enrollees and births with a LOS over 4 days. High performing MSAs in both groups also had more Medicaid patients, more non-profit and teaching hospitals, more hospitals that provide obstetric care, more hospital beds per 1,000, a greater number of FQHCs, and a higher density of primary care physicians. On average, the population in high performing MSAs was less likely to have received less than a high school or more than a college education. The population in high performing MSAs was also more likely to be Black, and less likely to be uninsured or impoverished compared to low performing MSAs in both treatment and control groups. I anticipate that these differences are explained by inherent differences between large, metropolitan areas and small, micropolitan ones.

There is little evidence of differential changes in the sample after APII implementation. In general, gaps between groups declined. For example, the difference in percent of HMO enrollees, LOS over 4 days, as well as Medicaid share decreased. Average percent cost sharing and education level rose across all groups, while the uninsurance and unemployment rates dropped. Other maternal, healthcare, demographic, and economic characteristics evolved similarly over time. To strengthen the assessment of covariate balance in Table 4-1, I evaluated the propensity score weights to ensure that there were no systematic differences in characteristics between treatment and controls in the pre- and post- periods (Table 2-7). The standardized mean differences (SMDs) between treatment and control groups showed significant improvement in covariate balance. After constructing the propensity score weights, all but three covariates had SMDs below 0.25, with the majority of these variables under 0.10 (Austin & Stuart, 2015; Stuart et al., 2014, 2013; Zhang et al., 2019).

Among high performing MSAs, unadjusted overall spending per episode was lower in Arkansas relative to control states prior to APII implementation ($9,666 vs. $11,507) (Table 4-2). Overall spending increased in both treatment and control groups after APII implementation. In Arkansas, per episode spending rose to $10,023 and $11,453 in Arkansas after partial and full implementation of APII, respectively. In control states, overall spending in high performing MSAs increased to $13,138 in the partial implementation period and $13,749 in the full implementation period. Among low performing MSAs, this trend was also evident. In Arkansas, pre-APII spending was, on average, $7,877, and increased to $10,227 and $11,833 after partial and full implementation of APII, respectively. In control states, per episode spending was $10,150 prior to APII, and increased to \$12,837 and \$15,650 in the partial and full APII implementation periods, respectively. Trends in overall episode spending by year, across groups, can be observed in Figure 4-1. These patterns persisted across categorical spending.

Among low performing MSAs, unadjusted quality of care was higher in Arkansas relative to control states, for all but three measures (gestational diabetes screenings, low-risk c-sections, and receipt of timely postpartum care). After implementation of APII, quality of care was higher in Arkansas for all prenatal screening measures, and remained lower for intrapartum and postpartum measures. Among low performing MSAs, unadjusted quality of care prior to APII was lower in Arkansas relative to control states for five of the eight quality measures, including chlamydia, HIV, and hepatitis B screenings, low-risk c-sections, and receipt of timely postpartum care. Low performing MSAs had lower quality rates compared to high performing MSAs for all measures, except for low-risk c-sections. This finding is likely a function of equally weighting each quality measure in the composite performance score. In the post-APII period, quality of care among low performing MSAs in Arkansas surpassed high performing MSAs for gestational diabetes screening rate after full implementation (83.39% vs. 81.30%, respectively). Trends in quality of care by year and across groups can be observed in Figure 4-1.

4.3.2. Association Between Bundled Payments and Healthcare Spending

Analyses on the effect of APII on per episode spending showed 16.2% greater overall savings in high performing MSAs relative to low performing MSAs after partial implementation (Standard Error (SE): 0.059; p-value<0.01) (Table 4-2). However, there was no evidence of differential savings in high versus low performing MSAs after full implementation of APII, relative to the pre-policy period. Savings during the partial implementation period appeared to be distributed across intrapartum facility spending (log odds: -0.321; SE: 0.074; p-value<0.01) and postpartum spending (log odds: -1.073; SE: 0.542; p-value<0.05).

4.3.3. Association Between Bundled Payments and Perinatal Quality

Analyses on the effect of APII on prenatal quality showed significant variation between high and low performing MSAs after partial implementation of APII (Table 4-3). Two of the three measures linked to gainsharing showed significantly greater improvements in low performing MSAs compared to high performing MSAs, in Arkansas relative to control states. Chlamydia and HIV screenings increased 75.6% and 97.5% more, respectively, in low performing MSAs compared to high performing MSAs for episodes paid under APII (Chlamydia – SE: 0.329; p-value<0.05; HIV – SE: 0.356; p-value<0.01). Two of the three prenatal screenings that were not linked to gainsharing showed the opposite effect, as high performing MSAs improved quality to a greater extent than low performing MSAs. After partial implementation of APII, gestational diabetes screenings increased significantly in high performing MSAs compared to low performing MSAs (log odds: 0.889; SE: 0.365; p-value<0.01). After full implementation of APII, Hepatitis B screenings increased significantly among high performing MSAs compared to low performing MSAs, for episodes paid under APII (log odds: 1.068; SE: 0.218; p-value<0.01). No other quality measures showed differential changes in prenatal quality after full implementation compared to the pre-APII period. There was no variation in effects between in high and low performing MSAs for intrapartum and postpartum quality of care measures (Table 4-4).

4.3.4. Decomposing Price and Quantity Effects

Decomposition analyses showed that savings in high performing MSAs was attributable to declines in both intrapartum prices and quantity of care (Table 4-5). After partial implementation of APII, intrapartum prices and quantity decreased 28.5% and 31.9% more, respectively, in high performing MSAs compared to low performing MSAs under APII relative to control states (Price – SE: 0.062; p-value<0.01; Quantity – SE: 0.080; p-value<0.01). There were no differential changes in price or quantity between high and low performing MSAs after full implementation of

APII, suggesting that the approach for achieving savings between high and low performing MSAs varied in the partial implementation post-period only.

4.3.5. Robustness Checks

Results were robust to alternate specifications and samples (Appendix). This included repeating analyses without multiple group propensity score weights (Table 4-8) and with MSA-level composite performance as a continuous, rather than binary, measure (Table 4-10). I also tested alternate samples by re-running models with alternate comparison groups (Table 4-11), and with inclusion of episodes that are excluded under APII reimbursement (Table 4-9). Since results were consistent across each of these models, it suggests that inferences from the main models can be generalized more broadly.

4.4. Discussion

This article presents evidence on the association between geographic variation in baseline performance and healthcare spending and quality among providers paid under a mandatory, multi-payer bundled payment program for perinatal episodes in Arkansas. This paper contributes to understanding the ways in which mandatory participation in bundled payments relates to total medical spending and utilization across the healthcare system. I find 16.2% greater healthcare savings overall among episodes in high performing areas compared to low performing areas as part of the episode-based payment program in the initial implementation period. Savings among high performing areas were due to declines in both hospital prices and intrapartum utilization. I also find short-term improvements in quality among low performing areas relative to high performing areas for prenatal screening measures that are linked to reimbursement as part of the bundled payment program. However, these quality gains are offset by smaller improvements for prenatal screening measures not tied to the payment incentive, relative to high performing areas. The results were generally consistent across model specifications and patient populations.

The results on healthcare spending appear to be consistent with existing studies. Prior research on the effects of bundled payments on spending found that participation is associated with greater savings in programs with voluntary, rather than mandatory participation (Finkelstein et al., 2018). Among voluntary bundled payment programs, participating hospitals tend to be large, non-profit, teaching hospitals with high patient volume and integrated PAC facilities (Navathe et al., 2017). Within this self-selected group, the greatest spending reductions are concentrated among top performers at baseline (Navathe et al., 2018). This research suggests that high performing healthcare systems may be more likely to both participate in bundled payment programs, and achieve greater savings within a given episode. My study confirms this result by demonstrating that high performing geographic areas reduce overall spending to a greater extent than low performing areas under bundled payments, at least in the short-term. This finding adds to the literature by exploring variation in savings across the full provider population facing a mandatory bundled payment incentive, and it also decomposes the mechanism for these savings.

Potential reasons for why low performing areas experienced smaller declines in healthcare spending are multifaceted. The most plausible explanation is that high performing areas had higher spending at baseline, so providers in this group had more opportunity for cost savings. Savings could be achieved through multiple avenues, such as a reduction in expensive, elective procedures and hospital prices. A related explanation is that areas with low performing providers had fewer opportunity to reduce spending without compromising quality. These areas had lower rates of low-risk c-sections at baseline in both treatment (18.37% versus 25.72%) and control groups (13.22% versus 15.62%) and fewer patients with an extended length of stay. Thus, it is conceivable that rates of other low-value, elective procedures may have been relatively limited as well. Without widespread “low hanging fruit” to pursue savings, these areas may have had fewer pathways to achieve reductions in episode spending.

Further, it is possible that savings in low performing areas were counteracted by the increased provision of high quality services for specific patient populations. For example, the financial incentive may have encouraged obstetricians to recommend use of a midwife or doula for socially underserved birthing people, who may have previously been unaware of this option or experienced barriers to access (Darling et al., 2019). Doula and midwifery assistance is associated with better birth outcomes via emotional, physical, and informational support, including lower likelihood of maternal complications and/or having a low birthweight baby (K. J. Gruber, Cupito, & Dobson, 2013). Analyses were unable to isolate the extent to which net savings were offset by an increase in the use of these safe, hospital-based childbirth services for vulnerable patients. If this mechanism were present, it would suggest that physicians responded to the financial incentive as intended, by improving care coordination over the entire episode.

Differential cost savings between high and low performing areas were not sustained beyond the first year of program implementation. A myriad of factors may have contributed. One possible reason is that high performing areas may have experienced a spillover from Medicaid to commercial enrollees after partial program implementation, while low performing areas did not. This explanation is plausible, since high performing areas had higher Medicaid enrollment (20.08% versus 13.13%), increasing the likelihood that physicians extended care practices across payer populations. Another explanation is the delayed onset of cost savings by providers in low performing areas. Further delineation for why differences in cost savings between high and low performing areas dissipated in the long-term warrants careful scrutiny.

This study provides greater context on the relationship between bundled payments and quality of care, especially as prior studies have shown mixed effects. Among low performing areas, I find greater rates of quality improvement for prenatal measures that are linked directly to the episode-based payment, and smaller rates of quality improvement for prenatal measures that do not directly contribute to reimbursement under the bundled payment design. This suggests that the financial incentive can be effective in encouraging low performing providers to offer better quality of care throughout the episode, but these improvements may be achieved at the expense of other important measures. It is unclear why low performing areas lag behind high performing areas for measures that are not tied to gainsharing. This variation could be due to the limited ability for low performers to prioritize a wide set of measures, in which case it makes sense to focus on measures that have payment implications. This result could also be driven by clinical complexity of patient populations or screening measures. Areas with high performing providers tend to have a higher proportion of Black patients, which, given current racial disparities in maternal care, complicates this narrative. Consistent with spending outcomes, effects of the episode-based payment on quality are not sustained beyond the first year of implementation. It is important for future work to address the factors contributing to this finding.

If results can be generalized, then these findings suggest that mandatory bundled payments may disadvantage low performers at baseline. This has several implications for efforts to expand bundled payment reforms. In terms of quality, low performers experience smaller quality improvement for measures not tied to reimbursement and greater gains for measures linked to payment, which suggests that linking measures to payment may be an effective strategy for bolstering quality among low performers at baseline, through the entire episode and across healthcare settings. Implications for spending are more complex because it is unclear whether modest cost savings among low performers at baseline are due to access-improving increases in utilization for their patients (e.g. increased prenatal testing, higher midwife and doula access, more comprehensive hospital care). Thus, it is important to assess the factors contributing to changes in spending over time. For example, if smaller healthcare savings are driven by the increased provision of services that improve patient outcomes, then it may be worthwhile for future bundled payment design to focus predominantly on unnecessary spending, rather than total spending.

Analyses had several limitations. The primary limitation was that the study may have been underpowered to detect policy-relevant differences in healthcare spending and quality for the high and low performing MSAs in the APII group compared to the controls. Part of this was related to the inherently high variability in the outcomes, which necessitated a large sample to detect significant differences. Another contributor was using Arkansas as the treatment state, which prompted concerns about whether the results of APII are generalizable to other geographic areas. Arkansas is a small, relatively rural state, and the MarketScan database only includes geographic indicators for micro and metropolitan areas. Thus, analyses did not include rural episodes, which may have biased results.

Some additional limitations to the analyses were related to the quality of the data. First, I could not identify individual payers or physicians in the MarketScan database. This blurred the distinction between physician and broader health system behavior changes and limited the ability to attribute the observed effects to physician behavior. It will be useful for future work to address this gap and examine these effects among individual physicians. Other limitations were related to the external validity of these findings. I focused on perinatal care, which, by definition, only applies to women. Only a segment of the population was eligible to enroll, thereby limiting the generalizability of my findings to this setting. Perinatal care has characteristics that reduce the likelihood for effects to stem from confounders, which may increase the extent to which results apply to other settings (e.g. it has high variation in costs and quality, as well as practical, elective treatment substitutes across delivery methods). These are factors that tend to be associated with supply-side, rather than demand-side, responses to financial incentives.

Finally, from a methods perspective, the DDD design rests on the assumption that no unobserved factors contribute to the observed effect. I addressed this concern by using a validated propensity score weighting approach, supplemented with several robustness checks. However, this approach does not account for time-varying, unobserved factors, which may weaken interpretation of results.

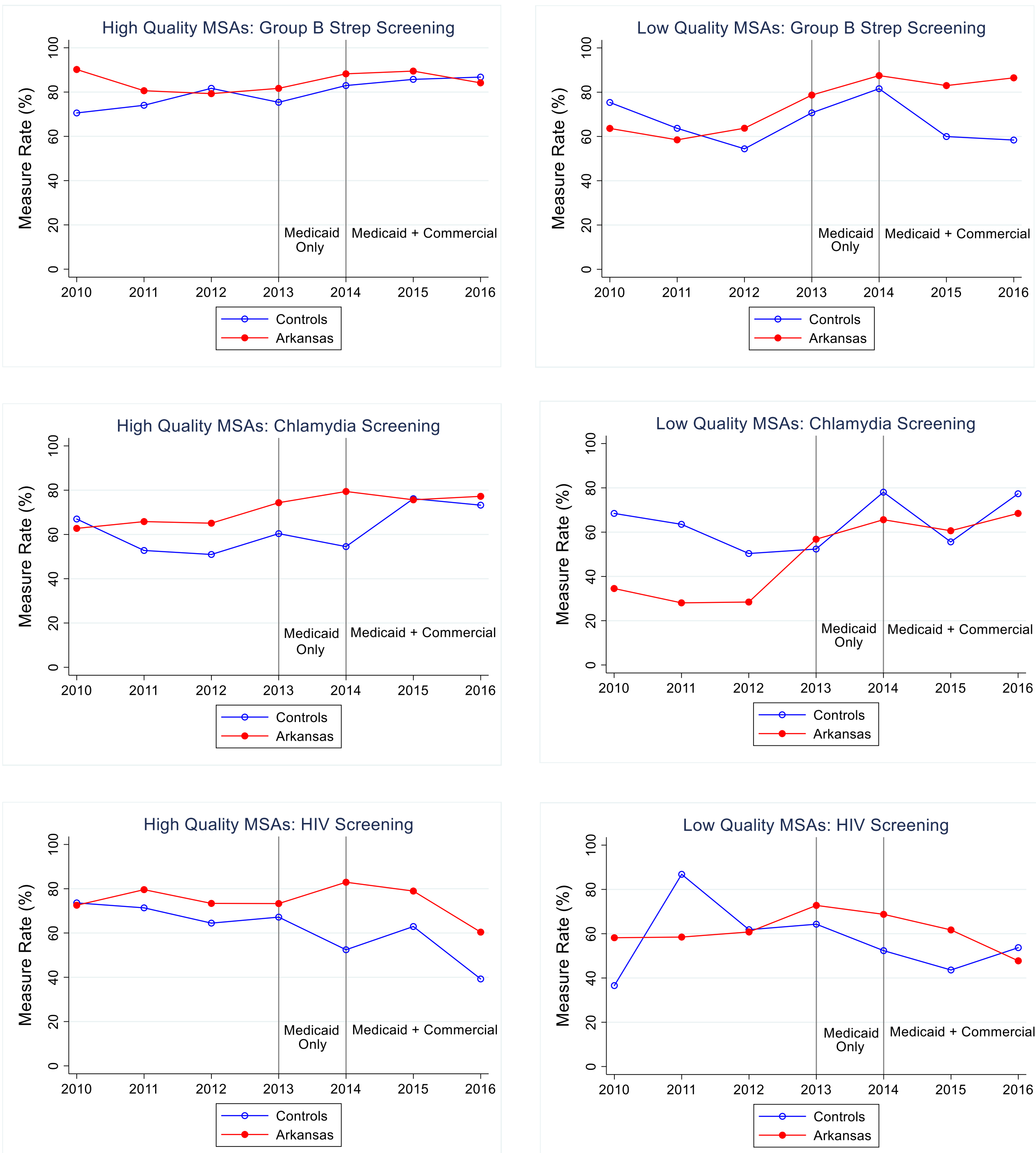
With high and rising healthcare spending in the U.S., generating savings through non-FFS payment reforms, like bundled payments, remains a priority, but there are uncertainties on best practices for incentive design. One such question is whether participating in bundled payment programs should be mandatory or voluntary, if high performing providers systematically choose to participate. This study presents evidence that under mandatory episode-based payments, geographic areas with high performing providers at baseline generate significantly greater savings through reduced hospital prices and volume of services. Geographic areas with low performing providers at baseline achieve larger improvement in quality measures directly linked to payment. However, neither of these differential effects are sustained in the long-term. The results underscore the need to delve further into factors that drive savings in low performing areas, and whether there is room to further reduce spending without compromising quality. The results are specific to perinatal care, but findings offer guidance on bundled payment design and participation for the general provider population. Continuing to explore variation in the effects of bundled payments across patient and provider populations is crucial to address through future research and policy.

4.5. Figures

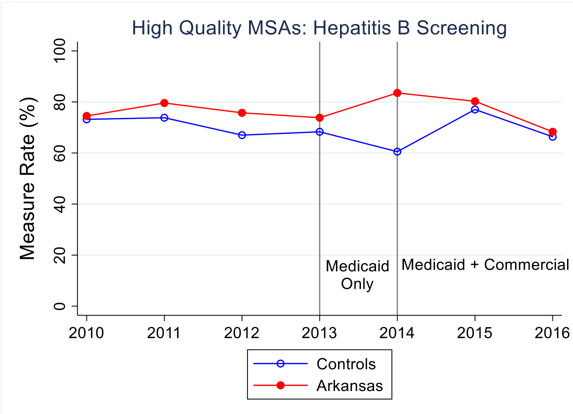
Figure 4-1 Trends in Quality for High versus Low Performing Metropolitan Statistical Areas at Baseline

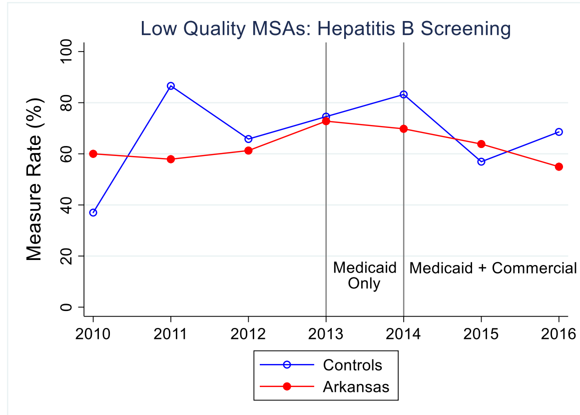
<u>Prenatal Screenings</u>

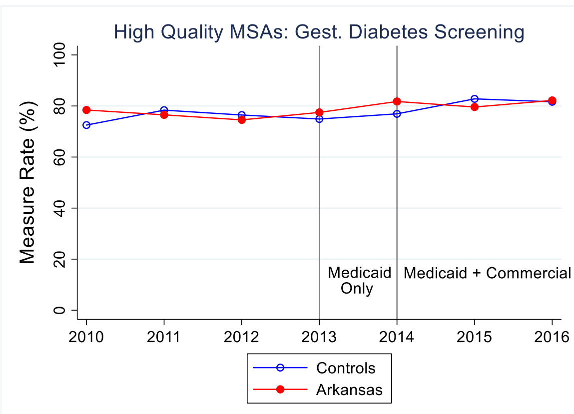
Linked to Gainsharing:

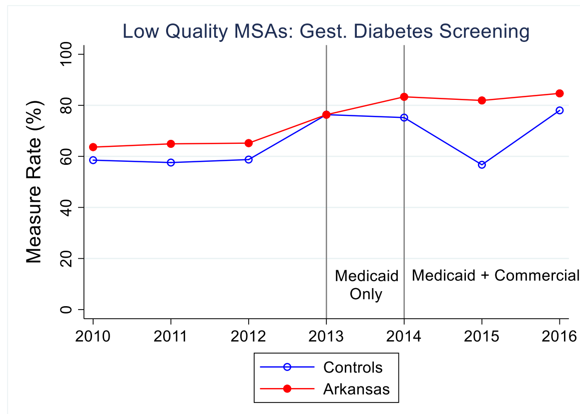


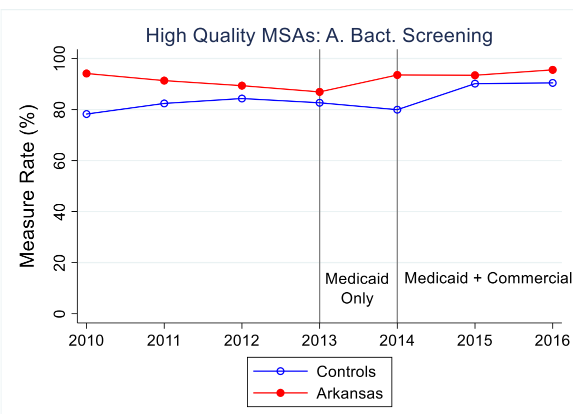
Not Linked to Gainsharing:

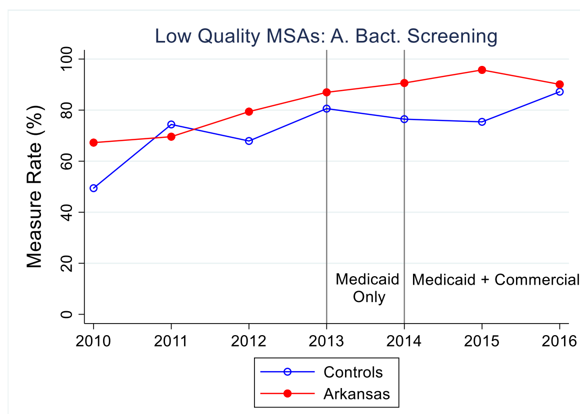




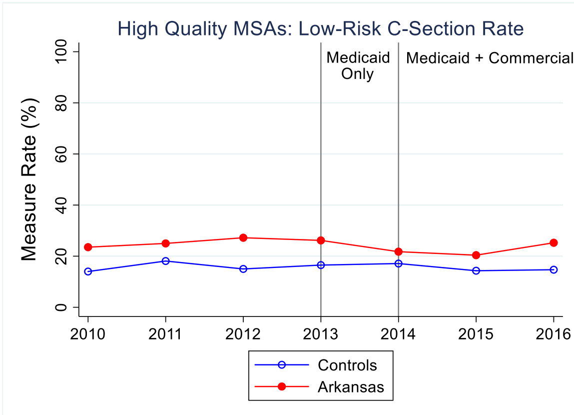


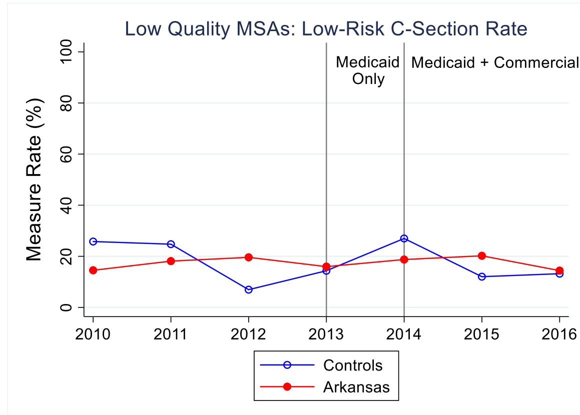




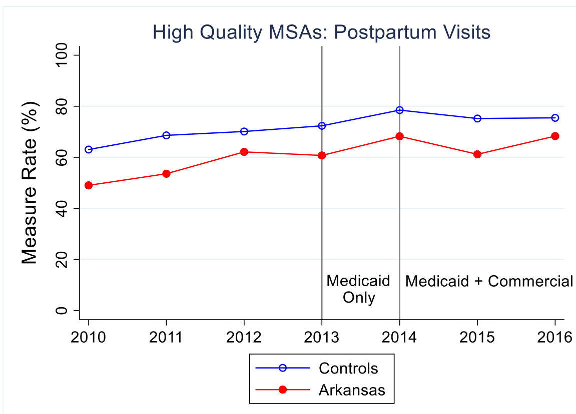


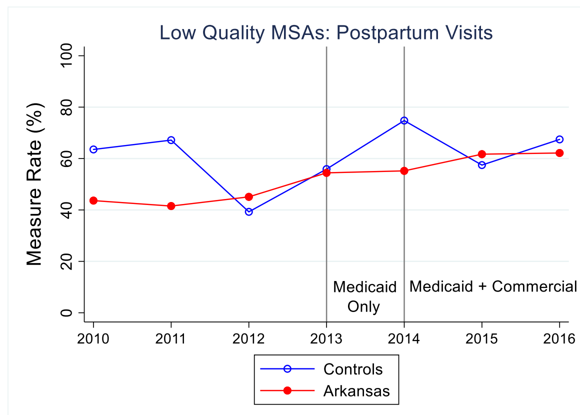
Perinatal Care:



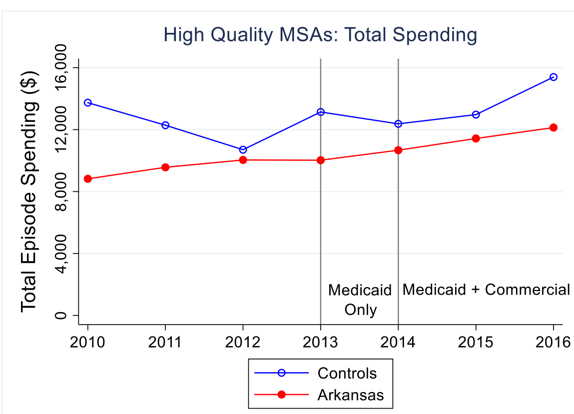


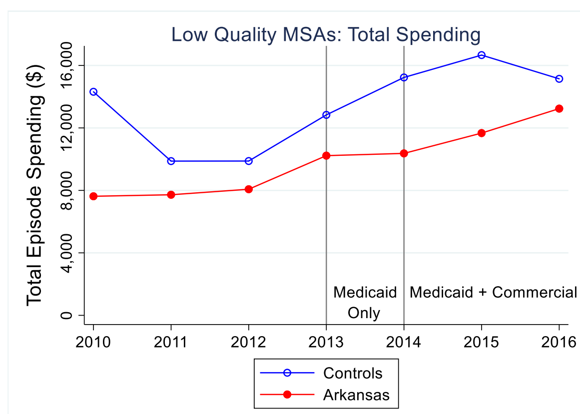
Postpartum Care:





Total Episode Spending:





4.6. Tables

Table 4-1 Summary Statistics of Arkansas and Control States, Pre and Post Arkansas Payment Improvement Initiative: High Performing versus Low Performing Metropolitan Statistical Areas

	Pre APII (2010-2012)				**Post APII (2013-2016)**			
	Arkansas		Control States		Arkansas		Control States	
	High Perf.	Low Perf.	High Perf.	Low Perf.	High Perf.	Low Perf.	High Perf.	Low Perf.
Maternal Characteristics								
% Maternal Age 35+	11.54	13.26	12.01	10.94	11.33	15.11	13.91	10.82
% LOS > 4	1.68	0.70	1.35	0.98	1.82	1.06	2.09	1.61
Health Plan								
HMO	3.13	0.47	3.58	0.16	1.68	0.85	2.46	0.81
PPO	74.76	78.60	75.36	75.64	68.81	65.74	66.60	63.15
POS	12.74	4.88	12.01	12.71	11.33	6.38	9.41	5.08
High Deductible	9.38	16.05	9.04	11.48	18.18	27.02	21.53	30.96
% Cost Sharing	18.94	16.80	17.86	18.71	21.97	23.34	19.68	22.34
MSA-Level Healthcare Characteristics								
% Medicaid	20.08	13.13	17.15	11.07	16.86	13.13	17.48	12.56
% Hospitals Non-Profit	46.05	26.12	37.04	31.34	48.54	30.40	52.84	53.13
% Hospitals Provide OB Services	25.59	37.44	28.30	26.75	29.22	41.55	33.43	32.52
% Teaching Hospitals	10.06	0.00	8.88	0.01	10.90	0.00	11.15	2.08
# of FQHCs	33.44	13.03	124.44	1.67	39.85	40.62	246.05	7.04
Beds per 1,000	5.49	2.78	3.18	2.04	5.60	2.92	1.73	0.59
PCPs per 1,000	0.77	0.64	0.89	0.57	0.80	0.68	0.90	1.75
MSA-Level Demographic Characteristics								
% < HS Education	12.13	14.95	11.65	14.48	10.82	14.46	10.68	14.36
% > College Education	26.88	28.04	33.05	23.34	28.23	28.58	36.16	23.07
% Population Black	19.82	6.30	16.41	9.13	20.94	7.48	18.57	9.56

MSA-Level Economic Characteristics								
% Uninsured	18.09	31.57	16.14	23.43	11.80	15.95	10.72	16.84
% Unemployed	7.13	6.14	7.54	6.28	5.39	4.86	5.57	5.78
% Poverty	15.40	18.35	14.68	19.87	15.15	17.08	13.87	17.12
Episodes	416	430	24,644	4,015	715	470	79,978	8,641
# of MSAs	3	3	125	26	3	3	125	26

Notes: Sample estimates are from the Truven MarketScan commercial claims database, using data from births during the study period. Healthcare characteristics are from the American Hospital Association Annual Survey, the Area Health Resource File, and the National Practitioner Data Bank. Demographic and economic characteristics are obtained from the U.S. Census American Community Survey

Table 4-2 Episode Spending: Variation in Effects of the Arkansas Payment Improvement Initiative: Variation Between Metropolitan Statistical Areas with High Performance versus Low Performance at Baseline

	Overall		Categorical Spending							
	Total Spending		Intrapartum Facility		Intrapartum Professional		Total Prenatal		Total Postpartum	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post										
High Perf.	-0.156**	-0.079*	-0.195**	-0.126**	0.050	-0.081	-0.184	0.042	-1.618	-0.560
	(0.072)	(0.045)	(0.078)	(0.057)	(0.077)	(0.050)	(0.151)	(0.122)	(1.003)	(0.527)
Low Perf.	0.007	-0.050	0.126***	-0.042	-0.048	0.176*	-0.051	-0.112	-0.544	-0.906
	(0.033)	(0.086)	(0.047)	(0.069)	(0.081)	(0.098)	(0.063)	(0.168)	(0.707)	(1.003)
Difference	-0.162***	-0.028	-0.321***	-0.084	0.098	-0.257**	-0.133	0.154	-1.073**	0.346
	(0.059)	(0.063)	(0.074)	(0.084)	(0.065)	(0.114)	(0.127)	(0.102)	(0.542)	(0.568)
N	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309
Dependent Variable Mean: %										
High Performing MSAs										
Pre-APII										
Treatment	9,666	9,666	4,559	4,559	2,918	2,918	1,765	1,765	424.30	424.30
Control	11,507	11,507	5,598	5,598	2,928	2,928	2,499	2,499	481.19	481.19
Post-APII										
Treatment	10,023	11,453	4,903	5,297	3,141	3,222	1,660	2,399	318.31	535.52
Control	13,138	13,749	6,229	6,400	3,430	3,712	2,855	3,129	625.06	508.56
Low Performing MSAs										
Pre-APII										
Treatment	7,877	7,877	3,991	3,991	2,171	2,171	1,501	1,501	214.26	214.26
Control	10,150	10,150	4,034	4,034	2,863	2,863	2,875	2,875	377.87	377.87
Post-APII										
Treatment	10,227	11,833	5,413	5,634	2,717	3,287	1,623	2,431	473.97	481.02
Control	12,837	15,650	5,154	7,383	3,765	3,425	3,393	3,875	525.33	967.66

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Spending models are estimated using a one-part generalized linear model with a log link and Gamma distributional family. Coefficients can be interpreted as the percent change in spending, pre- versus post-APII implementation. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-2012, prior to policy implementation being above a threshold of 68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 4-3 Prenatal Quality of Care: Variation in Effects of the Arkansas Payment Improvement Initiative: Variation Between Metropolitan Statistical Areas with High Performance versus Low Performance at Baseline

	Linked to Gainsharing						Not Linked to Gainsharing					
	Group B Strep		Chlamydia		HIV		Hepatitis B		Gestational Diabetes		A. Bact.	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post												
High Perf.	0.063	0.311	0.524**	0.461***	-0.311	0.758***	-0.149	0.166	0.440*	0.329*	-0.189	0.463
	(0.254)	(0.210)	(0.259)	(0.157)	(0.287)	(0.256)	(0.270)	(0.177)	(0.224)	(0.183)	(0.340)	(0.424)
Low Perf.	-0.162	0.516	1.280***	0.392	0.664***	0.769*	-0.071	-0.902***	-0.449**	0.151	-0.067	0.605*
	(0.299)	(0.363)	(0.253)	(0.339)	(0.252)	(0.463)	(0.208)	(0.286)	(0.219)	(0.280)	(0.255)	(0.310)
Difference	0.225	-0.205	-0.756**	0.069	-0.975***	-0.011	-0.078	1.068***	0.889***	0.177	-0.122	-0.143
	(0.346)	(0.349)	(0.329)	(0.343)	(0.356)	(0.385)	(0.312)	(0.218)	(0.365)	(0.274)	(0.309)	(0.360)
N	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309
Dependent Variable Mean: %												
High Performing MSAs												
Pre-APII												
Treatment	81.25	81.25	65.14	65.14	76.20	76.20	77.40	77.40	75.96	75.96	90.87	90.87
Control	78.29	78.29	53.62	53.62	67.41	67.41	69.51	69.51	76.38	76.38	83.00	83.00
Post-APII												
Treatment	81.68	87.02	74.35	77.48	73.30	73.09	73.82	76.72	77.49	81.30	86.91	94.27
Control	75.38	85.13	60.38	66.91	67.17	49.26	68.32	66.50	74.90	80.14	82.64	86.47
Low Performing MSAs												
Pre-APII												
Treatment	61.63	61.63	29.07	29.07	59.53	59.53	59.77	59.77	64.88	64.88	73.95	73.95
Control	58.36	58.36	55.24	55.24	67.45	67.45	70.00	70.00	58.40	58.40	68.64	68.64
Post-APII												
Treatment	78.70	85.71	56.80	65.12	72.78	58.80	72.78	62.46	76.33	83.39	86.98	92.03
Control	70.67	70.90	52.37	71.15	64.29	49.95	74.56	72.66	76.35	70.12	80.59	78.07

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-

2012, prior to policy implementation being above a threshold of 68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 4-4 Intrapartum and Postpartum Quality of Care: Variation in Effects of the Arkansas Payment Improvement Initiative: Variation Between Metropolitan Statistical Areas with High Performance versus Low Performance at Baseline

	Intrapartum Care		**Postpartum Care**	
	Low-Risk C-Sections		Postpartum Visit Within 6 Weeks	
	2013	2014-2016	2013	2014-2016
Treat * Post				
High Perf.	-0.816* (0.471)	-0.626** (0.291)	-0.576 (0.403)	-0.459* (0.242)
Low Perf.	-0.810*** (0.222)	-1.048* (0.557)	0.068 (0.153)	-0.186 (0.390)
Difference	-0.007 (0.363)	0.421 (0.436)	-0.644* (0.355)	-0.273 (0.279)
N	119,309	119,309	119,309	119,309
Dependent Variable Mean: %				
High Performing MSAs				
<u>Pre-APII</u>				
Treatment	25.72	25.72	56.49	56.49
Control	15.62	15.62	68.77	68.77
<u>Post-APII</u>				
Treatment	26.18	22.71	60.73	66.22
Control	16.52	15.52	72.34	76.53
Low Performing MSAs				
<u>Pre-APII</u>				
Treatment	18.37	18.37	43.49	43.49
Control	13.22	13.22	48.77	48.77
<u>Post-APII</u>				
Treatment	15.98	17.61	54.44	59.80
Control	14.35	20.01	55.90	68.24

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-2012, prior to policy implementation being above a threshold of 68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 4-5 Price-Quantity Decomposition of Intrapartum Facility Spending

	Intrapartum Facility Spending			
	Price Effect		Quantity Effect	
	2013	2014-2016	2013	2014-2016
Treat * Post				
High Perf.	-0.141** (0.068)	-0.090* (0.050)	-0.166* (0.094)	-0.164*** (0.058)
Low Perf.	0.145 (0.044)	-0.010 (0.051)	0.153*** (0.054)	-0.257*** (0.067)
Difference	-0.285*** (0.062)	-0.080 (0.068)	-0.319*** (0.080)	0.093 (0.077)
N	119,309	119,309	119,309	119,309

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Spending models are estimated using a one-part generalized linear model with a log link and Gamma distributional family. Coefficients can be interpreted as the percent change in spending, pre- versus post-APII implementation. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-2012, prior to policy implementation being above a threshold of 68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

4.7. References

Agarwal, B. R., Liao, J. M., Gupta, A., & Navathe, A. S. (2020). *The Impact Of Bundled Payment On Health Care Spending, Utilization, And Quality: A Systematic Review*. *1*(1), 50–57.

Agency for Healthcare Quality and Research. (2016). *AHRQ - Quality Indicators Technical Specifications*. Retrieved from https://www.qualityindicators.ahrq.gov/Modules/IQI_TechSpec_ICD10_v60.aspx

American Hospital Association. (2018). AHA Data Products: Annual Survey. Retrieved August 23, 2019, from https://www.aha.org/data-insights/aha-data-products

Arkansas Blue Cross Blue Shield. (2014). *Provider Manual: Payment Improvement Initiative*. Retrieved from www.ArkansasBlueCross.com.

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, *34*(28), 3661–3679. https://doi.org/10.1002/sim.6607

Barnett, M. L., Wilcock, A., McWilliams, J. M., Epstein, A. M., Joynt Maddox, K. E., Orav, E. J., … Mehrotra, A. (2019). Two-Year Evaluation of Mandatory Bundled Payments for Joint Replacement. *New England Journal of Medicine*, *380*(3), 252–262. https://doi.org/10.1056/NEJMsa1809010

Bhatt, S. P., Wells, J. M., Iyer, A. S., Kirkpatrick, D. P., Parekh, T. M., Leach, L. T., … Dransfield, M. T. (2017). Results of a Medicare Bundled Payments for care Improvement initiative for chronic obstructive pulmonary disease readmissions. *Annals of the American Thoracic Society*, *14*(5), 643–648. https://doi.org/10.1513/AnnalsATS.201610-775BC

Bolz, N. J., & Iorio, R. (2016). Bundled Payments: Our Experience at an Academic Medical Center. *Journal of Arthroplasty*, *31*(5), 932–935. https://doi.org/10.1016/j.arth.2016.01.055

Bronson, W. H., Kingery, M. T., Hutzler, L., Karia, R., Errico, T., Bosco, J., & Bendo, J. A. (2019). Lack of Cost Savings for Lumbar Spine Fusions after Bundled Payments for Care Improvement Initiative: A Consequence of Increased Case Complexity. *Spine*, *44*(4), 298–304. https://doi.org/10.1097/BRS.0000000000002812

Carroll, C., Chernew, M., Fendrick, A. M., Thompson, J., & Rose, S. (2018). Effects of episode-based payment on health care spending and utilization: Evidence from perinatal care in Arkansas. *Journal of Health Economics*, *61*, 47–62. https://doi.org/10.1016/j.jhealeco.2018.06.010

Chen, L. M., Ryan, A. M., Shih, T., Thumma, J. R., & Dimick, J. B. (2018). Medicare's Acute Care Episode Demonstration: Effects of Bundled Payments on Costs and Quality of Surgical Care. *Health Services Research*, *53*(2), 632–648. https://doi.org/10.1111/1475-6773.12681

Courtney, P. M., Ashley, B. S., Hume, E. L., & Kamath, A. F. (2016). Are Bundled Payments a Viable Reimbursement Model for Revision Total Joint Arthroplasty? *Clinical Orthopaedics and Related Research*, *474*(12), 2714–2721. https://doi.org/10.1007/s11999-016-4953-6

Curtin, B. M., Russell, R. D., & Odum, S. M. (2017). Bundled Payments for Care Improvement: Boom or Bust? *Journal of Arthroplasty*, *32*(10), 2931–2934. https://doi.org/10.1016/j.arth.2017.05.011

Darling, E. K., Grenier, L., Nussey, L., Murray-Davis, B., Hutton, E. K., & Vanstone, M. (2019). Access to midwifery care for people of low socio-economic status: A qualitative descriptive study. *BMC Pregnancy and Childbirth*, *19*(1), 416. https://doi.org/10.1186/s12884-019-2577-z

Dummit, L. A., Kahvecioglu, D., Marrufo, G., Rajkumar, R., Marshall, J., Tan, E., … Conway, P. H. (2016). Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes. *JAMA*, *316*(12), 1267. https://doi.org/10.1001/jama.2016.12717

Finkelstein, A., Ji, Y., Mahoney, N., & Skinner, J. (2018). Mandatory Medicare Bundled Payment Program for Lower Extremity Joint Replacement and Discharge to Institutional Postacute Care. *JAMA*, *320*(9), 892. https://doi.org/10.1001/jama.2018.12346

Glickman, A., Dinh, C., & Navathe, A. S. (2018). *The Current State of Bundled Payments*. Retrieved from https://ldi.upenn.edu/sites/default/files/pdf/LDI Issue Brief 2018 No. 3_41.pdf

Gray, C. F., Prieto, H. A., Duncan, A. T., & Parvataneni, H. K. (2018). Arthroplasty care redesign related to the Comprehensive Care for Joint Replacement model: results at a tertiary academic medical center. *Arthroplasty Today*, *4*(2), 221–226. https://doi.org/10.1016/j.artd.2018.02.002

Gruber, K. J., Cupito, S. H., & Dobson, C. F. (2013). Impact of Doulas on Healthy Birth Outcomes. *The Journal of Perinatal Education*, *22*(1), 49–58. https://doi.org/10.1891/1058-1243.22.1.49

HRSA. (2019). Area Health Resources Files. Retrieved September 11, 2019, from https://data.hrsa.gov/topics/health-workforce/ahrf

Johns Hopkins. (2016). Institute for Clinical and Translational Research: Truven Health MarketScan Research Databases. Retrieved September 11, 2019, from https://ictr.johnshopkins.edu/programs_resources/programs-resources/research-participant-recruitment-and-retention/truven-health-marketscan-research-databases/

Joynt Maddox, K. E., Orav, E. J., Zheng, J., & Epstein, A. M. (2018). Evaluation of Medicare's Bundled Payments Initiative for Medical Conditions. *New England Journal of Medicine*, *379*(3), 260–269. https://doi.org/10.1056/NEJMsa1801569

Jubelt, L. E., Goldfeld, K. S., Blecker, S. B., Chung, W. Y., Bendo, J. A., Bosco, J. A., … Horwitz, L. I. (2017). Early Lessons on Bundled Payment at an Academic Medical Center. *Journal of the American Academy of Orthopaedic Surgeons*, *25*(9), 654–663. https://doi.org/10.5435/JAAOS-D-16-00626

Kee, J. R., Edwards, P. K., & Barnes, C. L. (2017). Effect of Risk Acceptance for Bundled Care Payments on Clinical Outcomes in a High-Volume Total Joint Arthroplasty Practice After Implementation of a Standardized Clinical Pathway. *Journal of Arthroplasty*, *32*(8), 2332–2338. https://doi.org/10.1016/j.arth.2017.03.007

Markovitz, A. A., & Ryan, A. M. (2017). Pay-for-Performance: Disappointing Results or Masked Heterogeneity? *Medical Care Research and Review*, *74*(1), 3–78. https://doi.org/10.1177/1077558715619282

Martin, B. I., Lurie, J. D., Farrokhi, F. R., McGuire, K. J., & Mirza, S. K. (2018). Early Effects of Medicare's Bundled Payment for Care Improvement Program for Lumbar Fusion. *Spine*, *43*(10), 705–711. https://doi.org/10.1097/BRS.0000000000002404

National Practitioners Data Bank. (2019). The NPDB: Public Use Data File. Retrieved May 11, 2020, from https://www.npdb.hrsa.gov/resources/publicData.jsp

Navathe, A. S., Liao, J. M., Polsky, D., Shah, Y., Huang, Q., Zhu, J., … Emanuel, E. J. (2018). Comparison Of Hospitals Participating In Medicare's Voluntary And Mandatory Orthopedic Bundle Programs. *Health Affairs*, *37*(6), 854–863. https://doi.org/10.1377/hlthaff.2017.1358

Navathe, A. S., Troxel, A. B., Liao, J. M., Nan, N., Zhu, J., Zhong, W., & Emanuel, E. J. (2017). Cost of Joint Replacement Using Bundled Payment Models. *JAMA Internal Medicine*, *177*(2), 214. https://doi.org/10.1001/jamainternmed.2016.8263

Odum, S. M., Hamid, N., Van Doren, B. A., & Spector, L. R. (2018). Is there value in retrospective 90-day bundle payment models for shoulder arthroplasty procedures? *Journal of Shoulder and Elbow Surgery*, *27*(5), e149–e154. https://doi.org/10.1016/j.jse.2017.10.008

Siddiqi, A., White, P. B., Murphy, W., Terry, D., Murphy, S. B., & Talmo, C. (2018). Cost Savings in a Surgeon-Directed BPCI Program for Total Joint Arthroplasty - PubMed. *Surgeon Technology International*, *33*, 319–325. Retrieved from https://pubmed.ncbi.nlm.nih.gov/30029286/

Stuart, E. A., Huskamp, H. A., Duckworth, K., Simmons, J., Song, Z., Chernew, M., & Barry, C. L. (2014). Using propensity scores in difference-in-differences models to estimate the effects of a policy change. *Author Manuscript Health Serv Outcomes Res Methodol*, *14*(4), 166–182. https://doi.org/10.1007/s10742-014-0123-z

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8 SUPPL.8). https://doi.org/10.1016/j.jclinepi.2013.01.013

U.S. Census Bureau. (2018). *American Community Survey*. Retrieved from https://www.census.gov/programs-surveys/acs

Zhang, Z., Kim, H. J., Lonjon, G., & Zhu, Y. (2019). Balance diagnostics after propensity score matching. *Annals of Translational Medicine*, *7*(1), 16–16. https://doi.org/10.21037/atm.2018.12.10

4.8. Appendix Tables

Table 4-6 Tests for Equality Prior to Implementation of the Arkansas Payment Improvement Initiative: Quality and Spending Outcomes for High versus Low Performing Metropolitan Statistical Areas at Baseline

	Prenatal Screenings						**Intrapartum Care**	**Postpartum Care**	**Episode Spending**
	Linked to Gainsharing			Not Linked to Gainsharing					
	Group B Strep	Chlamydia	HIV	Hep. B	Gest. Diabetes	A. Bact.	Low-Risk C-Sections	Postpartum Care within 8 weeks	Total Spending
Treat * Quarter									
High Perf.	0.000	-0.021	0.105	0.065	-0.001	0.001	0.001	0.000	0.005 (0.006)
	(0.057)	(0.092)	(0.078)	(0.089)	(0.129)	(0.102)	(0.073)	(0.102)	
Low Perf.	0.000	-0.208)	-0.521	-0.616	0.001	0.000	0.000	-0.002	-0.022 (0.018)
	(0.190)	(0.150)	(0.471)	(0.544)	(0.545)	(0.536)	(0.205)	(0.410)	
Difference	0.000	0.187	0.626	0.681	-0.002	0.000	0.001	0.003 (0.401)	0.027 (0.020)
	(0.192)	(0.132)	(0.469)	(0.549)	(0.487)	(0.512)	(0.194)		
N	29,505	29,505	29,505	29,505	29,505	29,505	29,505	29,505	29,505
High Performing MSAs									
Pre-APII									
Treatment	81.25	65.14	76.20	77.40	75.96	90.87	25.72	56.49	9,666
Control	78.29	53.62	67.41	69.51	76.38	83.00	15.62	68.77	11,507
Low Performing MSAs									
Pre-APII									
Treatment	61.63	29.07	59.53	59.77	64.88	73.95	18.37	43.49	7,877
Control	58.36	55.24	67.45	70.00	58.40	68.64	13.22	48.77	10,150

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include pre-policy treatment-quarter trend coefficients with standard errors in parentheses. Spending models are estimated using a one-part generalized linear model with a log link and Gamma distributional family and coefficients can be interpreted as the percent change in spending, per quarter, pre-APII implementation. Quality models are estimated using multiple logistic regression. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-2012, prior to policy implementation being above a threshold of

68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 4-7 Covariate Balance Before and After Applying Multiple Group Propensity Score Weights

	Pre-APII (2010-2012)						**Post-APII (2013-2016)**					
	Initial Balance			*Propensity Score Balance*			*Initial Balance*			*Propensity Score Balance*		
	Treatment	Control	SMD	Treatment	Control	SMD	Treatment	Control	SMD	Treatment	Control	SMD
Maternal Characteristics												
Maternal Age 35+	0.09	0.13	-0.15	0.09	0.09	-0.01	0.09	0.14	-0.15	0.09	0.09	-0.01
LOS > 4	0.01	0.03	-0.10	0.01	0.01	0.00	0.02	0.03	-0.10	0.02	0.02	-0.02
Health Plan												
HMO	0.02	0.30	-0.85	0.02	0.02	0.00	0.01	0.17	-0.57	0.01	0.02	-0.01
PPO	0.77	0.53	0.51	0.77	0.76	0.03	0.68	0.54	0.28	0.68	0.65	0.06
POS	0.09	0.08	0.04	0.09	0.12	-0.13	0.09	0.05	0.17	0.09	0.07	0.08
High Deductible	0.13	0.09	0.12	0.13	0.10	0.08	0.22	0.24	-0.04	0.22	0.26	-0.11
% Cost Sharing	0.64	0.40	0.49	0.64	0.64	0.00	0.79	0.51	-0.07	0.79	0.80	-0.03
MSA-Level Healthcare Characteristics												
% Hospitals Non-Profit	0.05	0.40	-0.93	0.05	0.04	0.02	0.26	0.40	-0.31	0.26	0.40	-0.30
% Hospitals Provide OB Services	0.13	0.53	-0.96	0.13	0.08	0.11	0.12	0.49	-0.86	0.12	0.19	-0.14
% Teaching Hospitals	0.38	0.53	-0.31	0.38	0.23	0.30	0.49	0.46	0.06	0.49	0.34	0.30
# FQHCs	0.46	0.85	-0.89	0.46	0.46	0.00	0.76	0.84	-0.21	0.76	0.56	0.50
Beds per 1,000	1.00	0.47	1.51	1.00	0.83	0.48	1.00	0.51	1.39	1.00	0.61	1.10
PCPs per 1,000	0.50	0.53	-0.05				0.64	0.60	0.09	0.64	0.56	0.17
MSA-Level Demographic Characteristics												
% Less Than HS Education	0.65	0.74	-0.20	0.65	0.71	-0.12	0.48	0.51	-0.07	0.48	0.61	-0.27
% More Than College Education	0.00	0.37	-1.07	0.00	0.25	-0.74	0.38	0.66	-0.57	0.38	0.40	-0.03
MSA-Level Economic Characteristics												

% Uninsured	1.00	0.73	0.87	1.00	0.88	0.37	0.38	0.39	-0.03	0.38	0.35	0.06
% Unemployed	0.38	0.92	-1.37	0.38	0.43	-0.15	0.19	0.25	-0.15	0.19	0.16	0.07
% Poverty	0.62	0.55	0.15	0.62	0.67	-0.10	0.43	0.38	0.11	0.43	0.57	-0.28
% Population Black	0.44	0.36	0.17	0.44	0.37	0.15	0.58	0.41	0.35	0.58	0.48	0.21

Notes: Propensity scores were constructed using logistic regression. Covariates include binary indicators for each variable, coded as 1 if the value in a given MSA-Quarter is ≥ median and 0 if < median. Select covariates are excluded due to potential influence from APII in the post-period, including Medicaid share and OB malpractice payout. The model also included interaction terms for each of the covariates listed in Table 4-7 with an indicator for High Performing MSA, to increase covariate balance between High Low Performing MSAs. The estimate of interest is the Standardized Mean Difference (SMD), which provides an independent comparison between treated and control means.

Table 4-8 Sensitivity Analysis - Variation in Effects of the Arkansas Payment Improvement Initiative: Variation Between Metropolitan Statistical Areas with High Performance versus Low Performance at Baseline: Unweighted Analysis

	Prenatal Screenings: Linked to Gainsharing						Intrapartum Care	
	Group B Strep		Chlamydia		HIV		Low-Risk C-Section	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post								
High Perf.	-0.417**	-0.024	0.541***	0.504***	-0.109	0.599***	0.040	-0.100
	(0.179)	(0.181)	(0.141)	(0.092)	(0.187)	(0.123)	(0.094)	(0.081)
Low Perf.	-0.369	-0.326	0.823***	0.604**	0.144	-0.658**	0.019	0.262
	(0.447)	(0.385)	(0.301)	(0.291)	(0.296)	(0.312)	(0.113)	(0.203)
Difference	-0.049	0.303	-0.285	-0.100	-0.253	1.257***	0.021	-0.362*
	(0.463)	(0.405)	(0.318)	(0.284)	(0.326)	(0.304)	(0.142)	(0.212)
N	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309

	Prenatal Screenings: Not Linked to Gainsharing						Postpartum Care	
	Hepatitis B		Gestational Diabetes		Asympt. Bact.		Postpartum Visit within 8 Weeks	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post								
High Perf.	-0.134	0.209	0.069	0.303	-0.531***	0.362	0.201*	0.214***
	(0.164)	(0.139)	(0.199)	(0.187)	(0.144)	(0.349)	(0.113)	(0.076)
Low Perf.	0.187	-0.871***	-0.291	-0.315	-0.056	-0.019	0.492***	0.554***
	(0.325)	(0.289)	(0.389)	(0.332)	(0.394)	(0.250)	(0.152)	(0.178)
Difference	0.321	1.079***	0.360	0.618*	-0.474	0.381	-0.291*	-0.340*
	(0.346)	(0.301)	(0.416)	(0.365)	(0.396)	(0.414)	(0.168)	(0.176)
N	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309

	Episode Spending									
	Overall		**Categorical**							
	Total Spending		Intrapartum Facility		Intrapartum Professional		Total Prenatal		Total Postpartum	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post										
High Perf.	-0.041	-0.038	-0.009	-0.076*	-0.034	-0.017	-0.120***	0.041	-0.296***	0.008
	(0.027)	(0.030)	(0.042)	(0.041)	(0.027)	(0.030)	(0.038)	(0.049)	(0.074)	(0.069)
Low Perf.	0.170***	0.155***	0.151**	0.055	0.073	0.229***	0.176**	0.216***	0.747***	0.430***
	(0.049)	(0.055)	(0.063)	(0.066)	(0.058)	(0.073)	(0.073)	(0.069)	(0.164)	(0.105)

Difference	-0.211*** (0.055)	-0.193*** (0.050)	-0.160** (0.074)	-0.131** (0.056)	-0.106* (0.061)	-0.246*** (0.073)	-0.296*** (0.085)	-0.176** (0.083)	-1.043*** (0.168)	- 0.422*** (0.110)
N	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309

Notes: Sample estimates from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Spending models are estimated using a one-part generalized linear model with a log link and Gamma distributional family. Coefficients can be interpreted as the percent change in spending, pre- versus post-APII implementation. Quality models are estimated using multiple logistic regression. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-2012, prior to policy implementation being above a threshold of 68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table 4-9 Sensitivity Analysis - Variation in Effects of the Arkansas Payment Improvement Initiative: Variation Between Metropolitan Statistical Areas with High Performance versus Low Performance at Baseline: Inclusion of High-Risk Episodes

	Prenatal Screenings: Linked to Gainsharing						**Intrapartum Care**	
	Group B Strep		Chlamydia		HIV		Low-Risk C-Section	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post								
High Perf.	0.192	0.253	0.69***	0.717***	0.040	1.007***	-0.807**	-0.507**
	(0.236)	(0.190)	(0.220)	(0.128)	(0.254)	(0.227)	(0.401)	(0.249)
Low Perf.	-0.438**	0.453	1.634***	1.115***	0.394**	0.659*	-0.935***	-1.139**
	(0.196)	(0.314)	(0.217)	(0.334)	(0.160)	(0.362)	(0.264)	(0.545)
Difference	0.630***	-0.200	-0.937***	-0.398	-0.354	0.348	0.129	0.631
	(0.228)	(0.324)	(0.224)	(0.333)	(0.290)	(0.287)	(0.232)	(0.434)
N	167,403	167,403	167,403	167,403	167,403	167,403	167,403	167,403

	Prenatal Screenings: Not Linked to Gainsharing						**Postpartum Care**	
	Hepatitis B		Gestational Diabetes		Asympt. Bact.		Postpartum Visit within 6 Weeks	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post								
High Perf.	0.164	0.399**	0.750***	0.405**	0.064	0.567	-0.137	-0.312*
	(0.223)	(0.158)	(0.201)	(0.171)	(0.285)	(0.345)	(0.201)	(0.162)
Low Perf.	-0.111	-0.776***	-0.359**	0.352	-0.298	0.502*	0.173	0.096
	(0.161)	(0.243)	(0.159)	(0.278)	(0.286)	(0.263)	(0.119)	(0.295)
Difference	0.275	1.175***	1.109***	0.053	0.362	0.065	-0.310	-0.408*
	(0.262)	(0.210)	(0.225)	(0.284)	(0.287)	(0.291)	(0.190)	(0.241)
N	167,403	167,403	167,403	167,403	167,403	167,403	167,403	167,403

	Episode Spending									
	Overall		**Categorical**							
	Total Spending		Intrapartum Facility		Intrapartum Professional		Total Prenatal		Total Postpartum	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post										
High Perf.	0.090*	0.034	-0.045	-0.059	0.051	-0.067*	0.222**	0.217***	-0.046	-0.216
	(0.049)	(0.028)	(0.066)	(0.050)	(0.071)	(0.038)	(0.103)	(0.073)	(0.498)	(0.281)
Low Perf.	0.251***	0.166***	0.290***	0.079	-0.047	0.197***	0.314***	0.145*	0.038	-0.419
	(0.029)	(0.041)	(0.039)	(0.055)	(0.063)	(0.069)	(0.070)	(0.075)	(0.544)	(0.728)
Difference	-0.161***	-0.132***	-0.335***	-0.138**	0.098	-0.263***	-0.092	0.072	-0.083	0.203
	(0.051)	(0.045)	(0.050)	(0.069)	(0.060)	(0.080)	(0.113)	(0.079)	(0.439)	(0.550)

N	167,403	167,403	167,403	167,403	167,403	167,403	167,403	167,403	167,403	167,403

Notes: Sample estimates are from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-2012, prior to policy implementation being above a threshold of 68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 4-10 Sensitivity Analysis - Variation in Effects of the Arkansas Payment Improvement Initiative: Variation Between Metropolitan Statistical Areas with High Performance versus Low Performance at Baseline: Baseline Performance as Continuous Variable

	Prenatal Screenings: Linked to Gainsharing						**Intrapartum Care**	
	Group B Strep		Chlamydia		HIV		Low-Risk C-Section	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * Perf.	-1.095 (4.316)	-3.578 (4.166)	-9.445** (4.003)	-2.097 (4.374)	-12.479*** (4.740)	0.796 (3.975)	-0.680 (3.682)	4.539 (4.820)
N	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309

	Prenatal Screenings: Not Linked to Gainsharing						**Postpartum Care**	
	Hepatitis B		Gestational Diabetes		Asympt. Bact.		Postpartum Visit within 6 Weeks	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * Perf.	-4.366 (3.780)	10.724*** (2.579)	6.022* (3.143)	0.198 (3.642)	-6.463* (3.575)	-3.402 (4.826)	-6.853* (3.780)	-4.235 (2.578)
N	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309

	Episode Spending									
	Overall		**Categorical**							
	Total Spending		Intrapartum Facility		Intrapartum Professional		Total Prenatal		Total Postpartum	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * Perf.	-2.445*** (0.512)	-1.130 (0.713)	-3.509*** (0.790)	-1.375 (1.019)	-0.303 (0.809)	-3.621** (0.1.404)	-2.469** (1.062)	0.475 (1.083)	-14.805** (6.833)	0.331 (5.105)
N	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309	119,309

Notes: Sample estimates are from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-2012, prior to policy implementation being above a threshold of 68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 4-11 Sensitivity Analysis - Variation in Effects of the Arkansas Payment Improvement Initiative: Variation Between Metropolitan Statistical Areas with High Performance versus Low Performance at Baseline: Alternate Control Groups

	Prenatal Screenings: Linked to Gainsharing						**Intrapartum Care**	
	Group B Strep		Chlamydia		HIV		Low-Risk C-Section	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * [High – Low Quality]								
No CA	0.245	-0.112	-0.775**	0.214	-1.004***	0.031	-0.105	0.474
	(0.350)	(0.387)	(0.320)	(0.373)	(0.337)	(0.370)	(0.364)	(0.458)
No CT	0.216	-0.200	-0.751**	0.060	-0.907**	-0.036	-0.018	0.396
	(0.350)	(0.349)	(0.331)	(0.344)	(0.365)	(0.383)	(0.359)	(0.431)
No IA	0.225	-0.205	-0.756**	0.069	-0.975***	-0.011	-0.007	0.421
	(0.346)	(0.349)	(0.329)	(0.343)	(0.356)	(0.385)	(0.363)	(0.436)
No IN	0.450	-0.114	-0.739**	0.101	-1.240***	-0.344	0.384	0.880**
	(0.325)	(0.366)	(0.330)	(0.368)	(0.362)	(0.363)	(0.324)	(0.400)
No KY	0.061	-0.305	-0.732**	-0.031	-0.624*	0.010	0.035	0.311
	(0.365)	(0.410)	(0.322)	(0.374)	(0.332)	(0.385)	(0.400)	(0.426)
No LA	0.096	-0.401	-0.848***	-0.009	-0.960**	-0.155	-0.218	0.464
	(0.341)	(0.338)	(0.317)	(0.330)	(0.376)	(0.406)	(0.410)	(0.429)
No MI	0.174	-0.264	-0.764**	-0.001	-0.990***	-0.096	-0.043	0.407
	(0.348)	(0.342)	(0.327)	(0.341)	(0.350)	(0.379)	(0.385)	(0.440)
No MT	0.222	-0.208	-0.757**	0.067	-0.977***	-0.014	-0.004	0.425
	(0.346)	(0.348)	(0.328)	(0.343)	(0.356)	(0.384)	(0.364)	(0.436)
No NC	0.202	-0.264	-0.772**	-0.035	-1.061***	-0.125	0.017	0.500
	(0.377)	(0.365)	(0.367)	(0.318)	(0.369)	(0.363)	(0.361)	(0.436)
No ND	0.220	-0.208	-0.750**	0.075	-0.986***	-0.018	-0.025	0.415
	(0.348)	(0.349)	(0.332)	(0.343)	(0.356)	(0.384)	(0.371)	(0.436)
No NY	0.233	-0.209	-0.767**	0.076	-0.990***	0.000	-0.015	0.426
	(0.349)	(0.349)	(0.329)	(0.343)	(0.355)	(0.386)	(0.363)	(0.436)
No PA	0.406	-0.222	-0.843**	0.138	-1.064***	-0.187	-0.073	0.437
	(0.377)	(0.384)	(0.336)	(0.384)	(0.389)	(0.336)	(0.372)	(0.454)
No SC	0.243	-0.259	-0.728**	0.045	-0.979***	0.001	-0.103	0.384
	(0.367)	(0.370)	(0.344)	(0.353)	(0.358)	(0.387)	(0.378)	(0.433)
No TX	0.111	-0.482	-0.189	0.055	-0.367	1.309***	-0.061	-0.400
	(0.419)	(0.513)	(0.460)	(0.522)	(0.544)	(0.333)	(0.300)	(0.471)
No WV	0.162	-0.121	-0.760**	0.087	-0.888**	-0.114	0.006	0.442
	(0.336)	(0.360)	(0.329)	(0.355)	(0.364)	(0.363)	(0.361)	(0.450)
No WY	0.228	-0.207	-0.757**	0.067	-0.975***	-0.014	-0.006	0.424
	(0.247)	(0.350)	(0.329)	(0.344)	(0.356)	(0.386)	(0.363)	(0.437)

	Prenatal Screenings: Not Linked to Gainsharing						**Postpartum Care**	
	Hepatitis B		Gestational Diabetes		Asympt. Bact.		Postpartum Visit 6 Weeks	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * [High – Low Perf.]								
No CA	-0.142	1.009***	0.868***	0.147	-0.120	-0.242	-0.694*	-0.550**
	(0.318)	(0.241)	(0.282)	(0.323)	(0.328)	(0.403)	(0.361)	(0.278)
No CT	-0.044	1.060***	0.909***	0.189	-0.095	-0.142	-0.650*	-0.285
	(0.324)	(0.223)	(0.268)	(0.276)	(0.315)	(0.357)	(0.357)	(0.278)
No IA	-0.078	1.068***	0.889***	0.177	-0.122	-0.143	-0.644*	-0.273
	(0.312)	(0.218)	(0.265)	(0.274)	(0.309)	(0.360)	(0.355)	(0.279)
No IN	-0.150	0.998***	0.879***	0.185	0.155	-0.062	-0.439	-0.182
	(0.335)	(0.257)	(0.271)	(0.302)	(0.344)	(0.396)	(0.270)	(0.348)
No KY	0.114	1.105***	0.969***	0.257	-0.133	-0.319	-0.603	-0.355
	(0.316)	(0.241)	(0.254)	(0.301)	(0.322)	(0.376)	(0.391)	(0.268)
No LA	0.067	1.016***	1.080***	0.149	-0.125	-0.300	-0.669	-0.076
	(0.313)	(0.258)	(0.215)	(0.259)	(0.318)	(0.367)	(0.415)	(0.290)

No MI	-0.069	0.981***	0.920***	0.097	-0.079	-0.239	-0.674*	-0.272
	(0.310)	(0.214)	(0.266)	(0.265)	(0.310)	(0.350)	(0.363	(0.281)
No MT	-0.080	1.066***	0.888***	0.176	-0.126	-0.148	-0.645*	-0.275
	(0.312)	(0.218)	(0.264)	(0.274)	(0.310)	(0.360)	(0.354)	(0.280)
No NC	-0.076	0.947***	0.866***	-0.031	-0.116	-0.259	-0.651*	-0.140
	(0.331)	(0.240)	(0.259)	(0.271)	(0.315)	(0.352)	(0.362)	(0.295)
No ND	-0.083	1.063***	0.884***	0.172	-0.130	-0.150	-0.659*	-0.277
	(0.314)	(0.217)	(0.266)	(0.274)	(0.313)	(0.361)	(0.355)	(0.280)
No NY	-0.100	1.082***	0.885***	0.181	-0.131	-0.132	-0.649*	-0.270
	(0.312)	(0.221)	(0.266)	(0.274)	(0.311)	(0.362)	(0.354)	(0.279)
No PA	-0.075	0.974***	0.896***	0.209	0.004	-0.176	-0.824**	-0.100
	(0.362)	(0.223)	(0.293)	(0.271)	(0.348)	(0.374)	(0.353)	(0.284)
No SC	-0.030	1.120***	0.908***	0.172	-0.116	-0.121	-0.775**	-0.325
	(0.315)	(0.218)	(0.272)	(0.284)	(0.317)	(0.364)	(0.347)	(0.264)
No TX	-0.472	1.157***	0.341	0.169	-0.633	0.019	0.084	-0.663***
	(0.457)	(0.371)	(0.429)	(0.441)	(0.471)	(0.460)	(0.332)	(0.252)
No WV	0.007	0.954***	0.923***	0.148	-0.076	-0.170	-0.628*	-0.248
	(0.311)	(0.198)	(0.258)	(0.273)	(0.312)	(0.344)	(0.336)	(0.286)
No WY	-0.081	1.062***	0.886***	0.168	-0.126	-0.154	-0.647*	-0.265
	(0.312)	(0.219)	(0.263)	(0.273)	(0.309)	(0.359)	(0.355)	(0.278)

Episode Spending

	Overall		Categorical							
	Total Spending		Intrapartum Facility		Intrapartum Professional		Total Prenatal		Total Postpartum	
	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016	2013	2014-2016
Treat * Post * [High – Low Perf.]										
No CA	-0.172**	0.013	-0.354***	-0.025	0.124**	-0.252**	-0.115	0.243	-1.105**	0.335
	(0.067)	(0.060)	(0.082)	(0.083)	(0.063)	(0.109)	(0.141)	(0.098)	(0.540)	(0.677)
No CT	-0.174***	-0.026	-0.332***	-0.083	0.092	-0.261**	-0.145	0.159	-1.120**	0.373
	(0.060)	(0.062)	(0.074)	(0.083)	(0.068)	(0.113)	(0.132)	(0.102)	(0.564)	(0.562)
No IA	-0.162***	-0.028	-0.321***	-0.084	0.098	-0.257**	-0.133	0.154	-1.073**	0.346
	(0.059)	(0.063)	(0.074)	(0.084)	(0.065)	(0.114)	(0.127)	(0.102)	(0.542)	(0.568)
No IN	-0.120**	0.028	-0.293***	-0.060	0.098	-0.224*	-0.007	0.272***	-0.889*	0.621
	(0.057)	(0.054)	(0.068)	(0.085)	(0.062)	(0.117)	(0.126)	(0.081)	(0.510)	(0.733)
No KY	-0.143**	-0.008	-0.299***	-0.043	0.111	-0.246**	-0.121	0.103	-1.001*	0.268
	(0.069)	(0.065)	(0.080)	(0.086)	(0.070)	(0.119)	(0.141)	(0.093)	(0.542)	(0.536)
No LA	-0.176***	-0.022	-0.345***	-0.084	0.137**	-0.227**	-0.179	0.123	-1.151*	0.611
	(0.068)	(0.062)	(0.080)	(0.085)	(0.063)	(0.111)	(0.134)	(0.104)	(0.636)	(0.750)
No MI	0.168***	-0.030	-0.323***	-0.084	0.100	-0.260**	-0.143	0.148	-1.155**	0.333
	(0.062)	(0.063)	(0.076)	(0.085)	(0.063)	(0.116)	(0.133)	(0.102)	(0.566)	(0.583)
No MT	-0.162***	-0.028	-0.320***	-0.083	0.098	-0.256**	-0.132	0.154	-1.072**	0.345
	(0.059)	(0.063)	(0.074)	(0.084)	(0.065)	(0.114)	(0.127)	(0.102)	(0.540)	(0.570)
No NC	-0.154**	-0.043	-0.336***	-0.133*	0.123*	-0.259**	-0.096	0.160	-1.008*	0.472
	(0.060)	(0.061)	(0.073)	(0.079)	(0.067	(0.110)	(0.132)	(0.097)	(0.581)	(0.595)
No ND	-0.160***	-0.028	-0.323***	-0.084	0.102	-0.257**	-0.129	0.154	-1.068*	0.386
	(0.060)	(0.063)	(0.074)	(0.084)	(0.065)	(0.113)	(0.131)	(0.102)	(0.551)	(0.603)
No NY	-0.167***	-0.026	0.326***	-0.081	0.091	-0.254**	-0.136	0.156	-1.087**	0.349
	(0.059)	(0.063)	(0.074)	(0.084)	(0.067)	(0.112)	(0.126)	(0.103)	(0.545)	(0.568)
No PA	-0.194***	-0.034	-0.389***	-0.129	0.103	-0.286**	-0.159	0.202*	-1.034*	0.459
	(0.056)	(0.074)	(0.056)	(0.082)	(0.068)	(0.133)	(0.142)	(0.111)	(0.569)	(0.585)
No SC	-0.169***	-0.035	-0.335***	-0.102	0.095	-0.273**	-0.129	0.159	-1.131*	0.339
	(0.059)	(0.063)	(0.074)	(0.085)	(0.065)	(0.123)	(0.129)	(0.100)	(0.599)	(0.545)
No TX	-0.150**	-0.173**	-0.112	-0.041	-0.111*	-0.439***	-0.165	-0.085	-1.035**	-0.564
	(0.076)	(0.085)	(0.122)	(0.130)	(0.065)	(0.110)	(0.124)	(0.112)	(0.407)	(0.406)
No WV	-0.168***	-0.019	-0.319***	-0.085	0.091	-0.245**	-0.156	0.162	-1.026*	0.383
	(0.059)	(0.065)	(0.077)	(0.084)	(0.071)	(0.113)	(0.127)	(0.110)	(0.536)	(0.579)
No WY	-0.163***	-0.029	-0.321***	-0.084	0.098	-0.256**	-0.133	0.153	-1.078**	0.342
	(0.059)	(0.063)	(0.074)	(0.084)	(0.065)	(0.114)	(0.127)	(0.102)	(0.542)	(0.568)

Notes: Sample estimates are from the Truven MarketScan claims, using data from births between 2010-2016. Table cells include DDD coefficients with standard errors in parentheses. Coefficients are displayed as log odds. Standard errors are clustered at the State-MSA level. High Performing MSAs are determined by a composite performance score from 2010-2012, prior to policy implementation being above a threshold of 68.68%, which was the corresponding national average. Covariates include all variables in Table 4-1, plus mean malpractice payout, and quarter-year and state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

5. Conclusion

Contemporary policy debates are grappling with the design of physician payment incentives to balance the dual considerations of beneficiaries' access and quality of care, adequacy of payment rates to support provider costs and access to capital, and efficient use of healthcare services. Policy at the local, state, and federal levels have experimented with many different approaches. This dissertation examined examples of financial incentive policies to better understand the indirect, downstream effects on beneficiaries insured by other payers, racial minorities, and populations with low quality of care.

5.1. Summary of Chapters

The first chapter used passage of a 2014 and 2015 Medicaid payment reform in four states that discontinued payment for Early Elective Deliveries (EEDs), a low-value service in perinatal care, to evaluate the spillover effects on low-value quality measures in private insurance markets. The results showed a small, but significant 3.4% impact on decreasing rates of EEDs in the commercial sector. These spillover effects were present relative to a voluntary hard stop policy and a pay-for-performance bonus incentive, but not compared to a quality improvement initiative. There was also no evidence of physician-induced demand, under which the rate of more profitable services would increase in response to negative income shocks. These changes were the result of aligned financial objectives between physicians and for-profit hospitals, leading to positive impacts on healthcare quality for patients outside of the policy's target population.

The second chapter examined the effect of a 2013 multi-payer bundled payment program for perinatal episodes in Arkansas on place-based racial disparities in private insurance markets. The episode-based payment sought to shift financial risk to physicians by offering a single, risk-adjusted case rate for the entire course of care, rather than reimbursing providers separately for each individual service. Results showed disproportionate quality gains in areas with a high proportion of White patients in the short-term; however, areas with a high proportion of Black patients were able to close the gap in the long-term. This study highlights the role of physician payment on racial disparities, and suggests a need for financial incentives that are directly tied to racial equity.

The third chapter leveraged a multi-payer episode-based payment program in Arkansas with compulsory provider participation to test whether baseline quality was associated with divergent impacts on quality and spending. Findings demonstrated that geographic areas with high performing providers at baseline were able to generate 16.2% greater savings through reduced hospital prices and volume of services. However, areas with low performing providers at baseline achieved larger improvements in quality measures directly linked to payment. Analyses suggest that "average" effects observed in prior research masked heterogeneity across physicians. Future bundled payment programs with mandatory participation may experience challenges containing costs and improving quality among low performing providers.

Across these three studies, this dissertation suggests that financial incentives in physician payment are an important determinant of healthcare delivery for populations that were not directly targeted by the payment program. The first chapter showed that lowering the Medicaid price of a low-value service had welfare-improving effects in private insurance markets. It demonstrated that setting different fee schedules for high- and low-value services has potential to benefit commercial enrollees to a greater extent than other incentives, including voluntary non-financial incentives and financial bonuses.

The structure of financial incentives can also have differential implications for vulnerable populations. All physicians in Arkansas faced the same bundled payment incentive structure; however, as reflected in chapters two and three, physicians treating a high proportion of racial minorities, as well as those with lower quality at baseline, experienced delayed onset of quality improvement and cost savings compared to the average physician. Putting this into the context of the economic framework presented in the introduction, this dissertation has shown that physician agency varies across payers and patient populations, leading to significant differences in the amount and quality of medical care over time. As physician payment shifted towards value-based incentive structures, physicians adapted their behavior, even for populations not directly targeted by the payment change. This occurred across payers and disadvantaged populations. As future public policies are deliberated, it will be important to consider physician behavior in relation to all patient populations as the need for value continues to rise.

5.2. Takeaways and Policy Implications

The dissertation's implications for physician reimbursement policy and incentive design are nuanced. Overall, findings suggest that private and public payment reforms have potential to generate welfare-improving advances in healthcare spending and quality beyond the target population, but improvements are not equally experienced by all patients. In particular, my examination of multiple value-based payment reforms highlights the variation in positive spillovers from Medicaid to commercial markets across incentive structures and patients, the importance of tethering payment to specific quality metrics for vulnerable groups, and the potential tradeoffs between quality improvement and cost savings.

First, this work demonstrates the presence of positive spillovers from Medicaid payment incentives onto privately insured beneficiaries, across incentive designs and populations. I found that a Medicaid fee change discontinuing payment for a low-value service led to positive spillovers, but effects were small. In contrast, bundled payment reform led to large spillovers from Medicaid to commercial enrollees, with respect to cost savings and quality gains, but benefits were disproportionately felt by socially advantaged populations. While this dissertation has made strides towards understanding spillovers in the context of financial incentives, drivers of this phenomenon remain incompletely understood. Several complex factors contribute to quality improvement and cost containment. From an individual perspective, the patient-clinician interaction is crucial in determining treatment decisions. More broadly, this interaction is affected by health system characteristics, the culture of healthcare delivery, and individual patient and clinical factors. Deepening our understanding of the role of incentives and associated spillovers in achieving equitable health outcomes across diverse patients and providers is an exciting area for future research.

Second, it is important that quality measures are directly tied to physician reimbursement, across the continuum of care, especially for vulnerable patients. Achieving quality improvement for socially and medically disadvantaged patients is complex, and requires more than a simple financial incentive. Nonetheless, it is clear that financial incentives send a signal to physicians in regards to prioritization of quality measures and patient care goals, paving the way for more efficient gains among high-risk populations. This dissertation showed the range of physician responses to gainsharing measures across patient and provider characteristics, as well as incentive types. In response to a Medicaid reform discontinuing payment for a low-value service, reductions in overuse were concentrated among for-profit hospitals, underscoring the importance of shared financial objectives between physicians and hospitals. Further, physicians with more complex patient populations achieved greater initial success on dimensions of quality that were linked to payment, relative to measures that were not. Meanwhile, physicians with low-risk patients were able to realize quality gains across all measures, regardless of payment implications. Effectively, the financial incentive initiated early program success and avoided a period of worsening disparities for vulnerable populations and providers. Policymakers and insurers have faced measure selection challenges due to the added administrative burden and variation in what is considered "appropriate" treatment for different patient groups. This dissertation highlights the importance of concentrating on gainsharing measures most critical for high-risk populations, given finite resources.

Third, this dissertation illuminates the challenges of efforts to simultaneously contain costs and improve quality of care. Since financial incentives are often linked to both savings and quality, the physician must weigh the tradeoffs between saving healthcare dollars and improving quality, and in turn, deciding how to allocate efforts. In many cases, quality improvement requires increased healthcare spending which may further complicate decisions about allocation of effort. These decisions are influenced by an array of cultural and behavioral factors, and vary according to physician discretion, which is often immeasurable. Further, decisions vary across physicians' baseline performance; in my examination of the Arkansas bundled payment program, I found that physicians with high quality and low spending at baseline experienced slower efficiency gains, potentially because remaining improvements in performance are likely to be costly and more difficult (e.g. limited gains remain from "low hanging fruit"). Thus, physicians excelling in one area (e.g. high quality at baseline) may forego further progress in that space to concentrate on areas with more room for improvement, highlighting the imminent tradeoff between cost savings and quality. Future research should leverage qualitative data to assess the association between physician beliefs and attitudes on the response to financial incentives.

This dissertation contributes to a growing literature on financial incentives and physician behavior, with a focus on indirect effects of value-based payment reforms. Theoretically, value-based payment aligns incentives to maximize both profits and patient well-being, but socially and medically complex patients often require higher marginal effort to achieve these objectives. As a result, physicians may serve as especially sub-optimal agents for these populations, further raising concerns about inequity as payment reforms become more ubiquitous. This dissertation highlights the importance of assessing heterogeneity in direct and indirect effects of financial and non-financial incentives across providers in order to design performance improvement programs that account for these instances of imperfect physician agency and contribute to reaching the objective of healthcare value that is equitably experienced by all patients.

Curriculum Vitae

Name: Debra G. Bozzi

Education:

2016 – 2021	PhD in Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University (Expected May 2021) Concentration in Health Economics and Policy *Dissertation:* The Indirect Effects of Financial Incentives on Physician Behavior in Perinatal Care *Dissertation Committee:* Aditi Sen, PhD (Advisor); Darrell Gaskin, PhD; Elizabeth Stuart, PhD; Cynthia Minkovitz, MD (Chair); Craig Pollack, MD
2014 – 2016	MSPH in Health Policy, Johns Hopkins Bloomberg School of Public Health
2008 – 2012	B.A., Haverford College, *Major*: Political Science, *Minor*: Psychology *Independent Senior Thesis:* To What Extent Do Corporations Influence Federal Policymaking? An Analysis of the 2008 U.S. Farm Bill.

Positions and Employment:

2019 –	Research Assistant, Professor Craig Pollack, Johns Hopkins University
2018 – 2019	Research Assistant, Professor Bradley Herring, Johns Hopkins University
2017 – 2018	Research Assistant, Professor Aditi Sen, Johns Hopkins University
2015 – 2016	Provider Performance Program Analyst, Independence Blue Cross
2015	Health Policy Analyst, Maryland Department of Health and Mental Hygiene
2014 – 2015	Research Assistant, Professor Thomas LaVeist, Hopkins Center for Health Disparities Solutions, Johns Hopkins University
2012 – 2014	Senior Data Analyst, Pennsylvania Healthcare Quality Alliance
2012 – 2013	Research Assistant, Dr. Kristen Feemster, PolicyLab, Children's Hospital of Philadelphia

Honors and Awards:

2020	Winner, Delta Omega Honor Society Scholarship Award: Applied Research ($2,500 award), Johns Hopkins University
2020	Barbara Starfield Scholarship ($10,000 award), Johns Hopkins University
2018 – 2020	Health Policy & Management Department Tuition Scholarship, Johns Hopkins University
2016 – 2018	T32 National Research Service Award Pre-Doctoral Fellowship, Agency for Healthcare Research and Quality
2015 – 2016	Master's Tuition Scholarship (75% tuition award), Johns Hopkins University

Dissertation Papers:

1. **Bozzi DG.** The Indirect Effects of Medicaid Payment Changes: Evidence of Spillovers to the Commercially Insured *(Job Market Paper)*.

2. **Bozzi DG.** Place-Based Racial Disparities in Maternal Care: The Role of Episode-Based Payments

3. **Bozzi DG.** Do Effects of Mandatory Bundled Payments Vary by Baseline Quality? Evidence From Perinatal Care in Arkansas.

Publications:

Peer reviewed journal articles:

1. **Bozzi DG,** Nicholas LH. A Causal Estimate of Long-Term Healthcare Spending Attributable to Adult Body Mass Index. *Journal of Economics and Human Biology*. 2021; 41: 100985. Link.

Articles submitted for peer review:

1. Predmore Z, Doby B, **Bozzi DG**, Durand C, Sugarman J, Tobian A, Segev D, Wu AW**.** Experiences with and Attitudes towards the HOPE Act and HIV-Positive Organ Donation: A Qualitative Analysis of Barriers Experienced by Organ Procurement Organizations. (*Revise & Resubmit)*.

2. Ogunwole M, Karbeah K, **Bozzi DG**, Bower K, Cooper LA, Hardeman RR, Kozhimmanil KB**.** Health Equity Considerations in State Bills Related to Doula Care: A Review and Report Card. *(Under Review)*.

3. Pollack CE, **Bozzi DG**, Blackford AL, Deluca S, Thornton R, Herring B. Using the Moving To Opportunity Experiment to Investigate the Long-Term Impact of Neighborhoods on Healthcare Use by Specific Clinical Conditions and Type of Services. *(Under Review)*.

In Progress:

1. Pollack CE, **Bozzi DG**, Eisenberg MD, Matheson A, Laurent A. Exploring Impacts of Housing Assistance on Medicaid-Enrolled Children.

2. Keet CA, Matsui EC, Pollack CE, **Bozzi DG**, Peng R, Deluca S, Rule A, Wright R. The Effect of a Housing Mobility Program on Environmental Exposures, Asthma, and COVID-19.

3. Helms VE, Bachand JV, **Bozzi DG**. Effects of Smoke-Free Bans on Smoking Among Adults Receiving Federal Housing Assistance.

Non-peer-reviewed articles:

1. **Bozzi DG.** (2020). The Indirect Effects of Medicaid Payment Changes: Evidence of Spillovers to the Commercially Insured. 9th Annual Conference of the American Society of Health Economists, St. Louis MO (Virtual), 6/08/2020. Link.

2. Sen AP, **Gilbert (Bozzi) DG**, Asch D, Zhu J, Loewenstein G, Kullgren J, Volpp KG. (2018) The Effects of Financial Incentives on Intrinsic Motivation for Health Behaviors. 7th Annual Conference of the American Society of Health Economists, Atlanta GA, 6/10/2018. Link.

3. **Gilbert (Bozzi) DG**, Bigay K, Marlowe B, Muther E. (2017). Pennsylvania Healthcare Quality Alliance's State of the State Report. The Healthcare Improvement Foundation. Link.

4. Flynn K, Muther E, **Gilbert (Bozzi) DG**, Marlowe B. (2016). PRIDE: A Decade of Improving Healthcare Quality and Patient Safety in Southeastern Pennsylvania. The Healthcare Improvement Foundation. Link.

5. **Gilbert (Bozzi) DG**. (2012) To What Extent Do Corporations Influence Federal Policymaking? An Analysis of the 2008 U.S. Farm Bill. Haverford College Senior Thesis Archive. Link.

Presentations (as presenting author):

1. **Bozzi DG,** Pollack CE, Blackford AL, Herring B. Using Experimental Data from Moving To Opportunity to Investigate the Long-Term Impact of Neighborhoods on Hospital Admissions and ED Visits for Specific Clinical Conditions. Poster presentation, AcademyHealth Annual Research Meeting, Boston MA (Virtual), July 2020.

2. **Bozzi DG**. The Indirect Effects of Medicaid Payment Changes: Evidence of Spillovers to the Commercially Insured. Podium presentation, 9th Annual Conference of the American Society for Health Economists, St. Louis MO (Virtual), June 2020. [Presentation begins at 1:03:02].

3. **Bozzi DG**. Physician Agency. Class lecture, Health Economics, Johns Hopkins Bloomberg School of Public Health, Baltimore MD, Academic Years 2019-2020, 2020-2021.

4. **Bozzi DG.** Measuring Long-Term Healthcare Spending Attributable to Adult Obesity. Poster presentation, Annual Users Conference for the Panel Study of Income Dynamics, Ann Arbor MI, September 2019.

5. **Bozzi DG.** Measuring Long-Term Healthcare Spending Attributable to Adult Obesity. Poster presentation, 8th Annual Conference of the American Society for Health Economists, Washington DC, June 2019.

6. **Gilbert (Bozzi) DG**. Podium presentation, Agency for Healthcare Research and Quality 24th Annual NRSA Research Trainees Conference. Seattle WA, June 2018.

7. **Gilbert (Bozzi) DG**. Improving Health and Healthcare in the U.S.: The Role of Government. Panel moderator, Haverford College Health Economics Conference, Haverford PA, October 2017.

Professional Membership and University Services:

2018 –	American Society of Health Economists
2017 –	AcademyHealth
2018 –	American Economic Association
2017	Johns Hopkins University Krieger School of Arts and Sciences, Math Boot Camp for Economists
2017 – 2018	Johns Hopkins Bloomberg School of Public Health, Student Coordinating Committee Honors and Awards Chair
2015 – 2017	University of Pennsylvania Abramson Cancer Center, Philly Fights Cancer Young Friends Committee
2011 – 2012	Haverford College, Dining Services Development Committee
2008 – 2012	Women's Varsity Cross Country and Track

Teaching Experience:

2020 – 2021	Teaching Assistant, Introduction to the U.S. Healthcare System, Professors Daniel Polsky, Mariana Socal, and Craig Pollack
2017 – 2020	Teaching Assistant, Health Economics, Professor Aditi Sen
2018 – 2020	Teaching Assistant, Introduction to the U.S. Healthcare System, Professor Bradley Herring
2017 – 2020	The Tools of Public Health Practice, Professor Beth Resnick
2018 – 2020	Teaching Assistant, Fundamentals of Health Policy and Management, Professor Jon Vernick
2017 – 2020	Course Assistant, MSPH in Health Policy Seminar, Professor Beth Resnick
2019	Teaching Assistant, Introduction to Health Economics, Professor Doug Hough
2018, 2019	Teaching Assistant Behavioral Economics and Risk: Value-Based Payment Methods and Incentives, Professor Doug Hough
2018	Teaching Assistant, Introduction to Health Policy, Professor Gerard Anderson

Skills:

Computer/Statistical programs:
R, Stata, SAS, SQL, Teradata, ArcGIS, LaTeX, Microsoft Office, Mendeley, Lexus Research

Databases:
Truven MarketScan Commercial Claims and Encounters, Medicare Hospital Compare, Panel Study of Income Dynamics, Moving to Opportunity for Fair Housing Demonstration Survey, National Bureau of Economic Research Poverty Data, Medicaid Hospitalization and Emergency Department Claims, All-Payer Claims (CA, IL, MA, NY), Independence Blue Cross Claims and HEDIS Data, American Hospital Association Annual Survey, U.S. Census Bureau American Community Survey, Medical Expenditure Panel Survey, Healthcare Cost and Utilization Project, Health Resources and Services Administration Area Health Resource File, National Practitioners Data Bank